Journal of International Economics 144 (2023) 103793

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Full Length Articles

Within firm supply chains: Evidence from India☆

Shresth Garg [a,*], Pulak Ghosh [b], Brandon Joel Tan [a]

[a] Department of Economics, Harvard University, United States of America
[b] Indian Institute of Management Bangalore, India

ARTICLE INFO

Article history:
Received 30 June 2021
Received in revised form 29 June 2023
Accepted 30 June 2023
Available online 05 July 2023

Repository link: https://data.mendeley.com/datasets/czhpr6by9d/2

Keywords:
Vertical integration
Supply chains
Trade
Contracting

ABSTRACT

There are competing theories on whether vertical ownership is motivated by the transfer of physical inputs or the transfer of intangibles. Using administrative data on the universe of goods shipments in Karnataka, India, we show that the supply of goods along the production chain is an important rationale for vertical integration. First, we develop and estimate a gravity model of input sourcing, and find that: (1) establishments have a strong preference for sourcing their physical inputs from suppliers within the same firm relative to other frictions such as distance and state borders, and (2) the share of within-firm trade would be near 2% absent this preference for internal suppliers. Next, we compare this to the data and find that 38% of products are sourced by establishments exclusively from within the firm when a vertically integrated supplier exists; an order of magnitude higher than our 2% benchmark. Finally, we validate that within-firm sourcing is associated with determinants of physical supply chain transaction costs such as product specificity and R&D investment.

1. Introduction

Vertically integrated firms play a large and important role in the economy. Understanding the integration decision is a fundamental question in economics, going back at least to the seminal article by Coase (1937). There are competing theories in the literature on whether vertical ownership is motivated by efficiency gains from sourcing physical inputs from integrated upstream establishments, or the transfer of intangibles – such as organizational capabilities, know-how, managerial ability, and internal processes (Arrow, 1975; Teece, 1982; Gibbons, 2005).[1] In this paper, we offer a set of facts showing that the supply of physical goods along the production chain is an important rationale for vertical integration using administrative data on the universe of good shipments in Karnataka, India.

Tax records and survey data often track trade between firms, but rarely trade within firms. We exploit a recent policy change in India which generates transaction level data allowing us to map out the flow of physical goods both outside and within a firm. To improve tax compliance and reduce shipping times, firms in India are required to submit an electronic document to the

☆ We thank Enghin Atalay, Pol Antras, Dave Donaldson, Edward Glaeser, Nathan Hendren, Michael Kremer, Marc Melitz and Chad Syverson for helpful discussions and comments.

* Corresponding author.
E-mail addresses: garg@g.harvard.edu (S. Garg), pulak.ghosh@iimb.ac.in (P. Ghosh), btan@g.harvard.edu (B.J. Tan).

[1] See surveys by, Joskow (1985), P. G. Klein (2011), Shelanski and Klein (1995), and Lafontaine and Slade (2007).

https://doi.org/10.1016/j.jinteco.2023.103793

government prior to all movements of goods. For each establishment, we observe all inward shipments of inputs with information on product codes, product values, location of supplier, and whether the supplier is vertically integrated.[2] In our sample, establishments can potentially source 11% of their input value from other establishments owned by the same firm.[3]

First, we show that establishments have a strong preference for sourcing their physical inputs from suppliers within the same firm. We develop a gravity model of physical input sourcing, where establishments make sourcing decisions for inputs used in the production process. For each input, the establishment faces a set of suppliers, some of which may be vertically integrated. The decision to source from a supplier depends on whether the supplier is vertically integrated, the distance to the supplier, and whether the supplier is in the same state. The model yields a gravity specification which we estimate using our data. We find that firms exhibit a strong preference for sourcing from integrated suppliers and quantify that the magnitude of this preference is large relative to other frictions – three times larger than the preference for in-state suppliers and equivalent to a 95% reduction in distance.

Second, as a benchmark, we calculate that about 2% of products would be sourced from within the firm when an internal supplier exists, if vertical ownership is not considered. We use the estimated model to compute the share of within-firm trade that would exist if there is no preference for vertically integrated suppliers given geographic distance from intra-firm suppliers.

Third, we measure the observed share of within-firm trade in our data and find that downstream establishments source large shares of their inputs from their firms' upstream suppliers. In 38% of cases, establishments exclusively source from within the firm when they are able to, i.e. a vertically integrated supplier exists. Establishments rely on sources outside the firm in 58% of cases, and source from *both* within and outside the firm in 4% of the cases. The observed level of within-firm sourcing is an order of magnitude higher than our benchmark of 2% predicted by the model in the scenario with no preference for internal sourcing. Together, this indicates that vertically integrated physical chains are valuable to firms and is driving higher-than-expected levels of within firm trade.

Last, we validate that higher levels of within-firm sourcing are associated with determinants of physical supply chain transaction costs. We provide evidence that the preference for internal sourcing observed in our data is in line with theories of vertical integration as an efficient response to cost inefficiencies along the physical supply chain (Williamson, 1971, 1975, 1985; B. Klein et al., 1978). If a transaction generates ex-post rents, parties may engage in inefficient behavior to appropriate a larger share of the rents. Therefore, we should see higher levels of within-firm sourcing in situations where the problem of ex-post moral hazard, or holdup, is higher. Ex-post rents will be higher if the product is more specific to the production process and has fewer alternative buyers. Using Rauch (1999)'s classification, we confirm that within-firm sourcing is lower for products that are listed on exchanges. Listed products are typically standardized and have thick markets, reducing the potential for holdup. An alternative measure for the potential of holdup is the level of R&D requirement for the product. Using Nunn and Trefler (2013)'s measure of R&D intensity, we find that within-firm sourcing is higher for products that require more R&D expenditure.

This paper contributes to multiple strands of literature on vertical integration, which occupies a central role in organizational economics (Coase, 1937; Williamson, 1971; Bresnahan and Levin, 2012).

One set of theories place the rationale for vertical integration in the physical supply chain (Williamson, 1971, 1975, 1985; Goldberg, 1976; B. Klein et al., 1978; Joskow, 1987). Theoretical work in this tradition argues that certain features of physical transactions create particular problems for arms-length contracting, and offer explanations to why internal organization might lead to more efficient outcomes (Bresnahan and Levin, 2012). These can include problems of hold-up, uncertainty or contractual incompleteness, difficulty anticipating future contingencies, the need to use specific assets, and inability to measure and verify transaction outcomes. There are competing theories which place the rationale for integration in the transfer of intangibles. Theoretical work in this camp emphasizes the idea that firms may seek to expand or acquire other firms to leverage their internal capabilities or exploit superior management capabilities (Penrose, 1995; Cyert and March, 1966; Wernerfelt, 1984). These theories contend that organizations develop certain capabilities or know-how that is embodied by managers and employees, or in organizational routines, and these capabilities or knowledge cannot easily be traded or shared across firm boundaries (Malmgren, 1961; Cohen and Levinthal, 1990; Bresnahan and Levin, 2012).

Our paper builds on the many empirical studies that have attempted to test or apply contractual theories of integration (e.g. Monteverde and Teece, 1982; Masten, 1984; Walker and Weber, 1984; Masten et al., 1991; Crocker and Reynolds, 1993; Levin and Tadelis, 2010; Atalay et al., 2014; and Atalay et al., 2019).

We contribute to the empirical work by bringing in economy-wide transaction level data.[4] Our data allows us to measure input requirements at the establishment level and measure trade between establishments within the same firm. We also contribute to the literature by analyzing how within-firm trade varies across industries, products of different levels of standardization and R&D intensity, and a wide range of other correlates. Our paper is also unique in that we focus on empirically testing the drivers of vertical integration in a developing country context, where there may be higher contracting frictions (Boehm and Oberfield, 2018).

Understanding the reasons for vertical integration is critical from a policy and welfare perspective, as the benefits must be weighed against increased market power from consolidation (Salop and Scheffman, 1987; Krattenmaker and Salop, 1986; Hart and Moore, 1990; Ordover et al., 1990).

[2] An establishment is an economic unit owned by a firm and is distinct from other establishments based on the location of operation. For establishments that output multiple products, we treat vertical links for each product separately, and our main analysis is at the establishment-product level.

[3] 11% refers to the share of inputs (weighted by value) for which there is a vertically integrated supplier that the buying establishment can source from.

[4] In Appendix Section C, we conduct an aggregation exercise to demonstrate the importance of granular transaction-level data for analysis in our context.

It is important to highlight that the analysis in this paper has a few limitations. Our data captures only movement in physical goods and misses trade in services. Services form a significant part of the Indian economy, but we cannot measure within-firm trade in services in our data. Also, our data only captures formal firms that are registered under the tax regime. Registered firms have to report all movements of goods, even when transacting with an unregistered firm. Therefore our results are representative of the formal sector of the economy.

The remainder of the paper presents our results in more detail. In Section 2 we describe the data and variable construction. In Section 3 we present main results. We conclude in Section 4.

2. Data and variable construction

2.1. Data

In India, every registered business is required to submit an electronic document (known as an e-way bill)[5] to the government prior to any movement of goods valued above the threshold of Rs. 50,000 (~ $700 USD). This includes goods transported by road, air, railways, or water vessel. If the consigner is a registered taxpayer, they are responsible for generating an e-way bill. If they are not registered, then generating the e-way bill becomes the responsibility of the consignee or the person transporting goods. Notably, the bill is generated even if goods are shipped to a different establishment within the firm. The law was introduced to increase tax compliance and reduce shipping times. Government officials have the authority to intercept any conveyance to verify the e-way bill or the e-way bill number for all inter and intra-state shipments. The penalty for non-compliance is Rs 10,000 (~ $ 141 USD) or the value of tax-evaded, whichever is greater. In its first phase, the law covered only interstate shipments and in later phases was expanded to include intra-state shipments as well.[6]

We use administrative data on e-way bills from the state of Karnataka. Karnataka was the first state to require e-way bills at the intra-state shipment level, starting on April 1, 2018. Our dataset covers the universe of bills from April 1, 2018 to August 29, 2019. For each e-way bill, we observe the date of shipment, the tax ID (GSTIN) and ZIP code (PIN code) of the sender and the receiver, distance shipped, and the total value of the shipment.[7] A given shipment can contain multiple goods. For each good within a shipment, we observe its HS product code, its total value, and quantity. Firms report either 2, 4 or 8 digit HS product codes. For most of our analysis, we work with 4 digit codes, which refer to as a “product”. We also repeat our analysis by restricting to the subset of observations for which we see 8 digit codes for robustness. Our data includes all formal firms that ship goods in the state of Karnataka.[8] We provide some descriptive statistics in Tables 1 and A1. We observe over 86 million e-way bills (associated with over 196 million product shipments) from around 1.2 million firms. An average firm operates in 1.76 locations, is associated with 5 products, and makes 186 shipments in the period we observe. The average value of a shipment is around Rs. 200,000 (~ $ 2820 USD). The average total value of outward shipments for a given firm is Rs. 15 million (~ $ 210,000 USD). For each product code, we observe on average around 5500 establishments and 80,000 inward shipments.

2.2. Identifying firm establishments

For every shipment, for each party on both ends, we see the firm-level tax ID (GSTIN) and the location at the PIN code level.[9] We define an establishment as a GSTIN-PIN code pair. For example, if we see a firm shipping from 5 different PIN codes, we say that the firm has 5 establishments. Note that while we are not able to separate establishments that operate in the same PIN code, these are geographic areas that are very small and may effectively operate as a single entity.

2.3. Variable construction

We outline our baseline variable definitions in this section.

2.3.1. Identifying buyers and suppliers for each product

Each shipment made by a firm is associated with a unique e-way bill. The firm may include multiple products within a given shipment but has to report the product code (HS code) and value of each product. Firms may report 2, 4, or 8 digit product codes depending on firm size thresholds. For our main analysis, we remove observations at 2 digit level (which account for less than 0.4% of total value) and aggregate 8 digit codes to 4 digits. In the data, there are around 1300 four-digit codes and 10,600 eight-digit codes.

It is possible that an establishment both ships in and ships out a given product, for example, intermediaries specializing in trade. Excluding intermediaries that input and output the same product in our analysis is important. Consider a firm with a

---

[5] The filing of *E*-way bills was mandated with the rollout of Goods and Services Tax in India starting in 2017. There is a small but growing literature studying the impact of the tax regime. See for example Agarwal et al. (2019) and Leemput (2020).

[6] For more information refer to the information provided at https:/cleartax.in/s/eway-bill-gst-rules-compliance.

[7] We top-code all values at the 99th percentile.

[8] We note that we do not observe suppliers located outside the state of Karnataka who do not ship goods to Karnataka, even if they are vertically integrated with an establishment located in Karnataka. However, less than 5% of within-firm transactions take place across state lines. See Appendix Section B for a discussion.

[9] Each PIN code is mapped to exactly one delivery post office. There are over 150 thousand distinct PIN codes in India and it corresponds on average to an area of 21.22 km^2 and a population of roughly 8000 people.

Table 1
Descriptive statistics: firm and product level.

Statistic	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
Panel A: Firm Level					
Number of Locations	1.76	19.78	1	1	1
Number of Products	5.34	12.17	1	2	5
Number of Shipments	186.63	12,504.52	2.00	7.00	34.00
Average Value of Shipments	202,291.10	368,527.30	50,352.47	89,000.00	187,995.50
Total Value of Inward Shipments	15,820,502.00	1,016,284,634.00	50,000	339,871.5	2,283,888.0
Total Value of Outward Shipments	15,818,995.00	1,216,740,171.00	0	0	735,155.2
Panel B: Product Level					
Total Value of Inward Shipments	13,271,724,237.00	39,532,758,125.00	201,668,508.00	1,556,990,204.00	8,291,778,989.00
Total Value of Outward Shipments	13,271,724,237.00	39,532,758,125.00	201,668,508.00	1,556,990,204.00	8,291,778,989.00
Number of Firms	4824.37	8816.63	365.5	1541	5039
Number of Firm-Locations	6991.74	13,375.41	483.5	2077	7095
Number of Buying Firm-Locations	5917.28	11,710.90	340.5	1636	5875
Number of Selling Firm-Locations	1885.22	3489.62	164	613	1967
Number of In-Shipments	88,434.74	271,974.10	1636.00	11,951.00	63,445.00
Number of Out-Shipments	80,958.96	207,496.70	1618.00	11,743.00	59,994.50

Notes: Panel A provides descriptive statistics for the data aggregated to a firm level, where a firm is identified by a unique tax id (GSTIN). We report the average number of locations the firm operates in, the average number of product categories that the firm trades, the average number of shipments, the average value for each shipment and average value of inward and outward shipments. The numbers are reported in the local currency, INR. Panel B provides descriptive statistics aggregated to the product level. The product is defined as a 4 digit HS product code. Total Value of Inward and Outward Shipments is defined as the sum of all inward or outward shipment values for that product. Firms are identified by a unique tax id. Firm-locations are a tax id - ZIP code pair. We define buying and selling firm-locations according to Section 2. The Number of In/Out Shipments is the count of shipments for each product.

warehouse thats stores purchased inputs for distribution to the firm's production plants. Although this warehouse does ship inputs to another establishment within the same firm, such an establishment should not be classified as an upstream seller as it does not produce the input. Doing so would bias our estimates of within-firm shipping upwards.

To focus on cases where an establishment uses a product as part of the production process, we define an establishment as a consumer or a ‘net-buyer’ of a product if the total inward shipment value exceeds the total outward shipment value multiplied by a threshold. Similarly, we identify an establishment as a producer or a ‘net-seller’ of a product if the total outward shipment value of that product observed in our data exceeds the total inward shipment value multiplied by some threshold. For robustness, we confirm that our results are consistent across a range of threshold values.

More formally, consider a firm f with establishments $j_1, j_2, \cdots, j_n$. Total value of in-shipments of product p by establishment j_i is given by $\mathcal{I}^{j_i}(p)$ while total value of out-shipments of product p by establishment j_i is given by $\mathcal{O}^{j_i}(p)$. These values are zero if the establishment does not report a transaction in product p.

Consider the establishment j_i and the product p. It is a **net-seller** if $\mathcal{O}^{j_i}(p) > T_1 \times \mathcal{I}^{j_i}(p)$, where T_1 is a constant. It is a **net-buyer** if $\mathcal{I}^{j_i}(p) > T_1 \times \mathcal{O}^{j_i}(p)$. In our baseline $T_1 = 1.2$.

In Appendix Table A.2, we present descriptive statistics on buyers and sellers for different values of T_1, including the baseline $T_1 = 1.2$. For a given product, about 80% of the observed establishments are buyers and about 20% are suppliers. A firm on average sells 1.5 products and buys 7 products.

We also show that our results are robust to using an alternative method to identify intermediaries where we exclude establishments for which a large share of the products sold are products the establishment bought.

2.3.2. Measuring vertical integration

To measure potential for within-firm sourcing, we look at whether a downstream establishment has upstream sellers within the firm for an input product. As we observe all input purchases of an establishment, we can accurately determine if an establishment has integrated net-sellers for an input.[10] Note that in order to have an integrated upstream source, the net-seller does not have to sell to the net-buyer (our objective is to measure how often the net-seller does and does not sell to the net-buyer that is an establishment within the same firm).

The mere existence of an integrated upstream seller may not be enough to allow a firm to source from within, as the upstream seller's production scale may be too small relative to the downstream buyer's input requirements. External constraints may prevent the firm from reaching sufficient production capacity to meet its input requirements. For our baseline, we require that the total sales of all the integrated upstream establishments be at least 50% of downstream establishment demand. For robustness, we confirm that our results are consistent across a range of threshold values.

Following the previous notation, we say that establishment j_1 has a vertically integrated supplier if, $\sum_{k \in \{2,3,\cdots,n\}} O^{j_k}(p) > T_2 \times I^{j_1}(p)$, where T_2 is a constant equal to 0.5 in our baseline.

[10] We only consider downstream establishments located within the state of Karnataka, as we see their complete transaction details.

To an extent, the size of production is an endogenous decision of the firm, though limited by production scale constraints in the short-run. Nonetheless, for robustness, we also drop this restriction (by setting $T_2 = 0$) to test the sensitivity of our results.

2.3.3. Measuring utilization of within-firm production chains

Next, we define variables to measure the extent to which vertically integrated firms utilize their within-firm direct production links. We define a shipment i to be internal if both the sender and receiver on the e-way bill have the same unique tax ID (GSTIN).

Consider an establishment j, purchasing a product p for which there is an upstream integrated supplier. We compute the Within-Share$_{jp}$ as the value-weighted share of internal shipments in all inward shipments.

$$\text{Within-Share}_{jp} = \frac{\sum_{i \in \chi(j)} \text{Value}_{ip} * \mathbb{1}\{\text{Sender}_i \in f\}}{\mathcal{I}^j(p)}$$

where Value$_{ip}$ is the value of product p in shipment i, Sender$_i$ is the sending establishment, $\chi(j)$ is the set of shipments to establishment j, and $\mathcal{I}^j(p) = \sum_{i \in \chi(j)} \text{Value}_{ip}$.

3. Results

We present the results in this section. First, we develop a gravity model of input sourcing and present the estimation results. Next, we measure the realized level of within-firm sourcing. Last, we present results on the association between within-firm sourcing and determinants of physical supply chain transaction costs.

3.1. Gravity model of input sourcing

First, we develop and estimate a gravity model of input sourcing. The model will allow us to quantify the relative importance of various factors which govern the sourcing decisions of an establishment, including the preference for within-firm suppliers of physical inputs. Additionally, the model allows us to estimate the level of within-firm sourcing that would occur if firms do not take into account whether suppliers are integrated in the firm in their sourcing decisions.

3.1.1. Model

The model is based on revealed preference. An establishment j can choose to source input p from any other establishment that is a net-seller of p. The cost of sourcing depends on the distance to the seller, whether the seller is within the state, and if the seller is integrated.[11] The cost minimization problem faced by establishment j for each shipment of product p is,

$$\min_k \left\{ \underbrace{\alpha_0 + \alpha_1 \log\left(\text{Distance}_{jk}\right) + \alpha_2 \mathbb{1}\left(\text{Within} - \text{State}_{jk}\right) + \alpha_3\, \mathbb{1}\left(\text{Within} - \text{Firm}_{jk}\right) + \delta_{kp}}_{=\text{Cost}_{kp}} + \epsilon_{jkp} \right\}$$

where $k \in K_p$ is the set of all suppliers of input p, Distance$_{jk}$ is the distance between establishment j and supplier k, $\mathbb{1}(\text{Within-Firm}_{jk})$ is an indicator function for whether the supplier k is owned by the same firm as establishment j, and $\mathbb{1}(\text{Within} - \text{State}_{jk})$ is an indicator for whether the supplier k is located within the same state as establishment j. δ_{kp} is a fixed effect for the supplier-product kp, and ϵ_{jkp} is an establishment-supplier-product specific idiosyncratic shock.

We assume that ϵ_{jkp} follows an EV1 distribution, yielding the following expression for the probability that establishment j sources input p from k.

$$\Pr\left(\text{Source}_{jp} = k\right) = \frac{\exp\left(\text{Cost}_{kp}\right)}{\sum_{l \in K_p} \exp\left(\text{Cost}_{lp}\right)} \tag{1}$$

where Cost$_{kp}$ is the deterministic part of the cost specification. The final estimating equation is,

$$\Pr\left(\text{Source}_{jp} = k\right) = \exp\left(\alpha_0 + \alpha_1 \log\left(\text{Distance}_{jk}\right) + \alpha_2 \mathbb{1}\left(\text{Within–State}_{jk}\right) + \right. \tag{2}$$

$$\left. \alpha_3 \mathbb{1}\left(\text{Within–Firm}_{jk}\right) + \delta_{kp} + \gamma_{jp}\right).$$

where γ_{jp} is a buyer-product fixed effect.

[11] We describe our procedure for computing distance in Appendix Section D.

We estimate the equation using a Poisson regression.

3.1.2. Estimation results

The baseline estimates are presented in Table 3.

As expected, firms prefer to source from establishments that are closer and establishments that are located within the same state. The elasticities of bilateral trade flows with respect to Distance and Within-State are −0.19 and 1.2 respectively. To interpret the coefficients, we consider the average establishment in our sample which has many potential suppliers for each of its inputs and a baseline probability of sourcing from any given supplying establishment at 0.02%. A one standard deviation reduction in distance from the supplying establishment increases the probability to 0.03%. Moving the supplying establishment within state increases the probability of sourcing to 0.07%.

We find that firms exhibit a particularly strong preference for sourcing from integrated suppliers. The preference for within-firm suppliers is three times larger than the preference for in-state suppliers. Being a vertically integrated supplier has the same effect on an establishment's sourcing decision as a 95% reduction in distance.

In Appendix Table A8, we present results for specifications where we interact the Within-Firm variable with the Distance and Within-State variables. We find that the interaction coefficient between vertical integration and distance is positive. This indicates that the impact vertical integration has on the sourcing decision is stronger for more distant locations. We find that the interaction coefficient between vertical integration and being within-state is positive. This indicates that the impact of vertical integration on the sourcing decision is stronger for within state sourcing.

We report the estimation results for different thresholds of T_1 and T_2, including dropping all restrictions on the data, in Appendix Table A9.

3.1.3. Benchmarking

Next, we use the estimated model to compute the expected share of within-firm trade if firms did not have a preference for internal suppliers in their physical input sourcing decisions.

To do so, we set the coefficient on Within-Firm to zero, $\alpha_3 = 0$, such that there is no preference for within-firm sourcing. We then compute the new costs of sourcing from each potential supplier in this "benchmark" scenario, $\hat{\text{Cost}}_{kp}$, which depend only on Distance and Within-State.

$$\hat{\text{Cost}}_{kp} = \hat{\alpha}_0 + \hat{\alpha}_1 \log\left(\text{Distance}_{jk}\right) + \hat{\alpha}_2 \mathbb{1}\left(\text{Within} - \text{State}_{jk}\right) + 0 * \mathbb{1}\left(\text{Within} - \text{Firm}_{jk}\right) + \delta_{kp}$$

where $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ and $\hat{\delta}_{kp}$ are from the estimated model.

We use the new costs, $\hat{\text{Cost}}_{kp}$, to predict new probabilities of sourcing from each supplier in the model, following Eq. 1, and calculate the expected probability that the firm sources from an integrated supplier, conditional on the existence of one, i.e. there exists a seller with $\mathbb{1}_{jk}(\text{Within-Firm}) = 1$ in the set of suppliers K_p.

We estimate that 2.2% of products would be sourced from within the firm when an internal supplier exists if vertical ownership is not considered (Table 2 Row 3 Column 3).[12] This corresponds to 1.6% of input value (Table 2 Row 4 Column 3).

In Table A7, we report the benchmark within-firm trade share for different thresholds of T_1 and T_2,[13] including dropping all restrictions on the data, for robustness. The share that establishments would source, absent a preference for integrated suppliers, ranges between 2% to 3%.

This benchmarking exercise is key to interpreting whether the observed share of within firm trade is high or low in the next subsection. For instance, if different establishments of a firm are geographically concentrated we would expect to observe high levels of within-firm sourcing, even absent any preference for internal sourcing.

3.2. Utilization of vertically integrated production chains

Next, we measure the observed share of inputs establishments source from within the firm.

We first note that that 11% of inputs can potentially be sourced internally, i.e. 11% of inputs can be sourced by downstream establishments that have an integrated upstream establishment that produces the input (Table 2 Row 1), and firms that can source at least one product from within account for 61% of total output value (Table 2 Row 2).[14]

We report our baseline results in Fig. 1. The figure plots the distribution of Within-Share$_{jp}$ (defined in Section 2.3) over all establishments in our data, i.e. share of input-value sourced from within the firm conditional on the existence of a potential upstream supplier. We find that most establishments choose to source a given product either entirely from within or from outside, with few firms doing both. 38% of products are sourced by establishments exclusively from within the firm when a

[12] Note that in this exercise we only consider the extensive margin and assume that firms either source completely from within the firm or completely from outside the firm. Thus, the exclusive within-firm sourcing share (Table 2 Row 3 Column 3) is identical to the unweighted within-firm sourcing share (Table 2 Row 5).

[13] Refer to Section 2.3 for an explanation of thresholds T_1 and T_2.

[14] If we only consider firms that operate in more than one location, or multi-establishment firms, 13% of inputs can be sourced from an integrated upstream supplier. Single establishment firms cannot source from within by definition. Therefore, the proportion of within-firm sourcing goes up when we restrict to the subset of multi-establishment firms.

Table 2
Main results.

	Benchmark	Data
	(no pref. for internal supplier)	*(w/ pref. for internal supplier)*
Potential Within-Firm Sourcing	0.11	
Vertically Integrated Firms Share	0.61	
Exclusive Within-Firm Sourcing	0.022	0.38
Mean Within-Firm Sourcing (Weighted)	0.016	0.30
Mean Within-Firm Sourcing (Unweighted)	0.022	0.40

Notes: This table reports the main results in Section 3. Row 1 reports the total share of trade that can be conducted within firm boundaries. Row 2 reports the share of output accounted for by firms that are vertically integrated. Row 3 reports the share of product-establishments that source exclusively from within the firm conditional on a vertically integrated upstream establishment existing. Row 4 reports the share of sourcing that takes place within the firm conditional on a vertically integrated upstream establishment existing, weighted by value. Row 5 reports the unweighted share of sourcing that takes place within the firm conditional on a vertically integrated upstream establishment existing. In Column 3, we report results as measured in the data. In Column 2, we report results as predicted by the estimated gravity model with the coefficient on Within-Firm equaling zero, $\alpha_3 = 0$ (see Subsection 3.1).

Table 3
Estimation of gravity model of input sourcing.

	(1)	(2)
Log (Distance) (α_1)	−0.192	−0.191
	(−64.26)	(−69.81)
Within-State (α_2)	1.208	1.226
	(84.29)	(88.06)
Within-Firm (α_3)	3.777	
	(184.93)	
Observations	68,926,278	68,926,278

t statistics in parentheses
Notes: This table reports estimates of Eq. 2 from a Poisson regression of the share of sourcing on distance, an indicator for same firm ownership, and an indicator for being in the same state. Each observation is a buyer-potential supplier-product. The model and variable definitions are outlined in Subsection 3.1.

Table 4
Product specificity and R&D investment: within-firm sourcing.

	Within-firm Sourcing					
	(1)	(2)	(3)	(4)	(5)	(6)
Listed on Exchange	−9.336	−7.517	−8.184			
	(0.220)	(0.127)	(0.220)			
Log R&D Intensity				7.215	4.430	8.054
				(0.379)	(0.220)	(0.405)
Supplier Scale Requirement	Yes	No	No	Yes	No	No
Intermediary Exclusion Conditions	Yes	No	No	Yes	No	No
HS Digits	4	4	8	4	4	8
N	1,049,568	2,407,013	1,218,983	1,023,023	2,355,140	1,193,647
R^2	0.162	0.091	0.033	0.157	0.089	0.031

Notes: This table reports estimates from a linear regression of the share of within firm sourcing on the product being listed on an exchange, or the level of R&D intensity (see Eqs. 3 and 4). Product section and district fixed effects are included. Columns (1) and (4) correspond to our baseline measurement methodology. In Columns (2), (3), (5) and (6), we drop all restrictions on supplier scale and conditions to exclude intermediaries. In Columns (1), (2), (4) and (5), we use 4 digit product codes. In Columns (3) and (6), we use 8 digit product codes. Details on the model and variable construction can be found in Section 3.3.

vertically integrated supplier exists (Table 2 Row 3 Column 3), 58% are sourced exclusively from outside the firm and the remaining 4% being sourced from *both* within and outside the firm.

The observed level of sourcing is an order of magnitude higher than our benchmark of 2.2% predicted by the model in a scenario where firms do not have a preference for internal sourcing. This indicates that vertically integrated physical chains are valuable to firms and are driving higher than expected levels of within firm trade.

We also report the weighted average of Within-Share$_{jp}$ over all establishments in our data, where we weigh by each establishment's total purchase value of product *p*. Thus, each number represents the share of total trade that takes place within firms, out of the total potential trade that can take place within the firm. In our preferred specification, the average is 30% (Table 2 Row 4 Column 3). We also report the unweighted average at 40% (Table 2 Row 5 Column 3).

Table 5
Product specificity and ex-ante investment: extensive margin.

	Vertically Integrated Supplier Exists					
	(1)	(2)	(3)	(4)	(5)	(6)
Listed on Exchange	−0.011	−0.012	−0.051			
	(0.001)	(0.001)	(0.001)			
Log R&D Intensity				0.076	0.091	0.045
				(0.001)	(0.001)	(0.001)
Supplier Scale Requirement	Yes	No	No	Yes	No	No
Intermediary Exclusion Conditions	Yes	No	No	Yes	No	No
HS Digits	4	4	8	4	4	8
N	8,783,388	8,783,388	4,832,176	8,594,585	8,594,585	4,771,606
R^2	0.024	0.023	0.012	0.025	0.023	0.011

Notes: This table reports estimates from a linear regression of an indicator for the existence of a vertically integrated supplier on the product being listed on an exchange, or the level of R&D intensity. Product section and district fixed effects are included. Columns (1) and (4) correspond to our baseline measurement methodology. In Columns (2), (3), (5) and (6), we drop all restrictions on supplier scale and conditions to exclude intermediaries. In Columns (1), (2), (4) and (5), we use 4 digit product codes. In Columns (3) and (6), we use 8 digit product codes. Details on the model and variable construction can be found in Section 3.3.

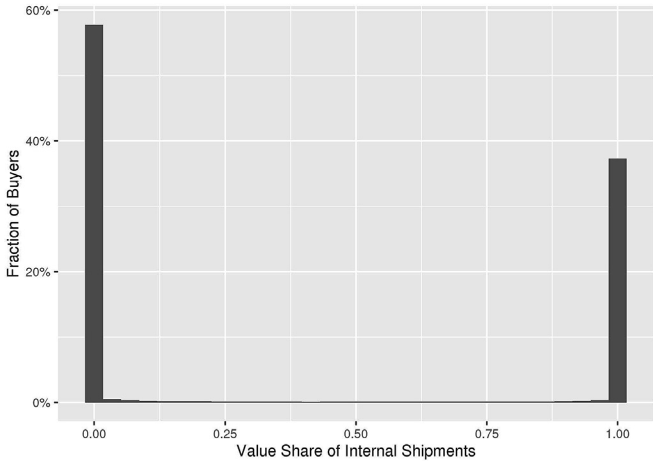

Fig. 1. Share of Within Firm Sourcing by Buyers.
Notes: This figure present a histogram of the share of within firm purchases made conditional on having a vertically integrated upstream establishment. We use our baseline measures to define vertical links (as explained in Section 2.3), i.e., $T_1 = 1.2$ *and* $T_2 = 0.5$.

In Appendix Section B, we report a range of robustness checks. First, we show that our results are similar for different thresholds of $T_1 \in \{1, 1.2, 1.5\}$ and $T_2 \in \{0.5, 1\}$.[15] We also present results without various restrictions discussed so far, as well as with alternative restrictions on the data. We find that across specifications the share of within trade along vertically integrated supply chains is large and significant. The proportion of establishments, sourcing exclusively from the integrated supplier, remains between 35% - 40%, compared to 38% in the baseline. Similarly, the weighted and unweighted proportion also remains between 30% - 40%. Finally, we also report additional results on the level of within-firm sourcing by product categories, and the share of production that is shipped to internal establishments from the upstream establishment's perspective.

3.3. Product specificity and R&D investment

Next, we validate that higher levels of within-firm sourcing is associated with the determinants of transaction costs in physical supply chains.

According to transaction cost theory, in settings with ex-post quasi rents, contracting parties in a supply chain face the problem of ex-post moral hazard. Parties may engage in inefficient behavior to appropriate a larger share of the quasi-rents (B. Klein et al., 1978). In such settings, vertical integration may be an efficient response. Therefore, we should see higher levels of within-firm sourcing in settings with higher quasi-rents.

[15] The thresholds T_1 and T_2 determine the set of establishment-products for which there is an upstream integrated supplier over which the mean of $\text{Within} - \text{Share}_{jp}$ is calculated. See Sectiom 2.3 for more details.

Quasi rents are higher if the value of alternative use is low and the product is relationship specific. We use the classification by Rauch (1999) to measure the importance of relationship-specific investment at the product level. This measure indicates whether or not a given input is sold on an organized exchange and whether or not it is reference-priced in a trade publication. If an input is sold on an organized exchange then the market for this input is thick, there are alternative buyers, and the quasi rents are low. Similarly, quasi rent might be lower for products that are referenced in trade publications. Trade publications are only produced if there is a sufficient number of purchasers of the product. Therefore, if an input is reference priced in a publication, then this indicates that there exists a reasonably large number of potential buyers and sellers of the input. Rauch's original classification uses the 4-digit SITC Rev. 2 system. Each industry is coded as being in one of the following three categories: sold on an exchange, reference priced, or neither. We use a 4-digit SITC to 10-digit HS concordance from Feenstra and Hanson (1996) to match the classification to our data. Rauch has both a liberal estimate and a conservative estimate. We use the liberal estimate and being listed on an exchange in our main specification, $1\left(\text{Listed on Exchange}_p\right)$.[16]

An alternative, albeit imperfect, measure of quasi-rents is the level of R&D investment required for a product. Products with higher levels of investment may be more specific to a relationship and therefore more subject to ex-post moral hazard.[17] To measure R&D intensity we use data from Nunn and Trefler (2013) who calculated R&D intensity based on data from the Bureau van Dijk's Orbis dataset. R&D Intensity is defined as R&D expenditures to total sales, and denoted $\text{R\&DIntensity}_p$ for product p. The data is provided at the HS 6 digit level, and we aggregate this measure to the HS 4 digit level by taking the mean over all HS 6 digit intensity measures that correspond to a given HS 4 digit code.

Our estimating equations are

$$\text{Within}-\text{Share}_{jp} = \beta_0 + \beta_1 \text{Log}\left(\text{R\&D Intensity}_p\right) + \text{Section}_p + \text{District}_j + e_{jp} \tag{3}$$

$$\text{Within}-\text{Share}_{jp} = \beta_0 + \beta_1 1\left(\text{Listed on Exchange}_p\right) + \text{Section}_p + \text{District}_j + e_{jp} \tag{4}$$

where Section_p is the product section of product p (see Table A6), District_j is the district where establishment j is located, and e_{jp} is an error term. Within-Share_{jp} is the share of product p that establishment j sources from within the firm, as defined in Section 2.3.

We first focus on the level of within-firm trade along vertically integrated links (the”intensive margin”).

The results are presented in Table 4. We find that a product that is listed on exchange is 9 percentage points less likely to be sourced from within. Doubling R&D requirement increases within-firm sourcing by 7 p.p. The results are quantitatively similar when we drop restrictions on supplier scale and include intermediaries, and when we use 8 digit product codes instead of 4 digits.

We find consistent results when we focus on the extensive margin, i.e. whether an establishment has a vertically integrated supplier. We present the results in Table 5. We find that a product that is listed on exchange is 1 percentage point less likely to have a vertically integrated supplier. Doubling R&D requirement increases within-firm sourcing by 8 p.p. The results are also quantitatively similar when we drop restrictions on supplier scale and conditions to exclude intermediaries, and when we use 8 digit product codes instead of 4 digits.

In Appendix Section E, we report additional results for other variables that may influence sourcing decisions. We find that both a greater number of and more frequent shipments increase the likelihood of sourcing from within the firm. Larger firms in terms of total value, number of establishments, and number of products are far more likely to take advantage of their integrated supply networks. Last, more competition upstream increases within-firm sourcing, while competition downstream reduces it.

4. Conclusion

This paper uses detailed administrative data on the movement of physical goods both within and outside the firm from Karnataka, India to offer a set of facts showing that the supply of physical goods along the production chain is an important rationale for vertical integration using administrative data on the universe of good shipments in Karnataka, India.

First, we develop a gravity model of physical input sourcing, where establishments make sourcing decisions for inputs used in the production process. The model yields a gravity specification which we estimate using our data. We find that firms exhibit a strong preference for sourcing from integrated suppliers and quantify that the magnitude of this preference is large relative to other frictions. As a benchmark, we also calculate that about 2% of products would be sourced from within the firm when an internal supplier exists if we shut down the preference for sourcing from within the firm.

Next, we measure the observed share of within-firm trade in our data and find that around 38% of products are sourced by establishments exclusively from within the firm when a vertically integrated supplier exists, an order of magnitude higher than our benchmark.

Last, we find that within-firm sourcing is higher for products with higher expected transaction costs in the physical supply chain. We show that products that require ex-ante investments or are more specific to the production process are more likely to be sourced within the firm.

[16] Our results are similar when using the conservative estimate.
[17] The measure is imperfect for our use as there may be products, say computers, that have high R&D intensity but low specificity.

Data availability

Within-Firm Supply Chains: Evidence from India Replication Package (Original data) (Mendeley Data)

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jinteco.2023.103793.

References

Agarwal, N., Ashlagi, I., Azevedo, C.R., Featherstone, Eduardo, Karaduman, Ö., 2019. Market Failure in Kidney Exchange. NBER Working Paper Series. No. 24775 pp. 1–52.
Arrow, K.J., 1975. Vertical integration and communication. Bell J. Econ. 6.
Atalay, E., Hortacsu, A., Syverson, C., 2014. Vertical integration and input flows. Am. Econ. Rev. 104 (4), 1120–1148.
Atalay, E., Hortaçsu, A., Li, M.J., Syverson, C., 2019. How wide is the firm border? Q. J. Econ. 134 (4), 1845–1882.
Boehm, J., Oberfield, E., 2018. Misallocation in the Market for Inputs: Enforcement and the Organization of Production. *National Bureau of Economic ResearchWorking Paper Series*. No. 24937.
Bresnahan, T.F., Levin, J.D., 2012. Vertical Integration and Market Structure. (Working Paper No. 17889)National Bureau of Economic Research.
Coase, R.H., 1937. The nature of the firm. Economica 4 (16), 386–405.
Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: a new perspective on learning and innovation. Adm. Sci. Q. 35.
Crocker, K.J., Reynolds, K.J., 1993. The efficiency of incomplete contracts: an empirical analysis of air force engine procurement. RAND J. Econ. 24.
Cyert, R.M., March, J.G., 1966. A Behavioral Theory of the Firm. by Richard M. Cyert and james g. March. Vol. 60. Cambridge University Press (CUP).
Feenstra, R.C., Hanson, G.H., 1996. Globalization, outsourcing, and wage inequality. Am. Econ. Rev. 86 (2), 240–245.
Gibbons, R., 2005. Four formal(izable) theories of the firm? J. Econ. Behav. Organ. 58.
Goldberg, V.P., 1976. Regulation and administered contracts. Bell J. Econ. 7.
Hart, O., Moore, J., 1990. Property Rights and the Nature of the Firm (tech. rep. No. 6).
Joskow, P.L., 1985. Vertical integration and long-term contracts: the case of coal-burning electric generating plants. J. Law Econ. Org. 1.
Joskow, P.L., 1987. Contract duration and relationship-specific investments: empirical evidence from coal markets [Publisher: American Economic Association]. Am. Econ. Rev. 77 (1), 168–185.
Klein, Peter G., The Make-or-Buy Decision: Lessons from Empirical Studies (April 2004). Available at SSRN: https://ssrn.com/abstract=529962 or http://dx.doi.org/10.2139/ssrn.529962
Klein, B., Crawford, R.G., Alchian, A.A., 1978. Vertical integration, appropriable rents, and the competitive contracting process. J. Law Econ. 21 (2), 297–326.
Krattenmaker, T.G., Salop, S.C., 1986. Competition and cooperation in the market for exclusionary rights. Am. Econ. Rev. 76.
Lafontaine, F., Slade, M., 2007. Vertical integration and firm boundaries: the evidence. J. Econ. Lit. 45 (3), 629–685.
Leemput, E.V., 2020. A Passage to India: Quantifying Internal and External Barriers to Trade.
Levin, J., Tadelis, S., 2010. Contracting for government services: theory and evidence from u.s. cities*. J. Ind. Econ. 58 (3), 507–541.
Malmgren, H.B., 1961. Information, expectations and the theory of the firm. Q. J. Econ. 75.
Masten, S.E., 1984. The organization of production: evidence from the aerospace industry. *The*. J. Law Econ. 27.
Masten, S.E., Meehan, J.W., Snyder, E.A., 1991. The costs of organization. J. Law Econ. Org. 7.
Monteverde, K., Teece, D.J., 1982. Supplier switching costs and vertical integration in the automobile industry. Bell J. Econ. 13.
Nunn, N., Trefler, D., 2013. Incomplete contracts and the boundaries of the multinational firm. J. Econ. Behav. Organ. 94, 330–344.
Ordover, J.A., Saloner, G., Salop, S.C., 1990. American economic association equilibrium vertical foreclosure equilibrium vertical foreclosure. Source: Am. Econ. Rev. 80.
Penrose, E., 1995. The Theory of the Growth of the Firm. Oxford University Press.
Rauch, J.E., 1999. Networks versus markets in international trade. J. Int. Econ. 48 (1), 7–35.
Salop, S.C., Scheffman, D.T., 1987. Cost-raising strategies. J. Ind. Econ. 36.
Shelanski, H.A., Klein, P.G., 1995. Empirical research in transaction cost economics: a review and assessment. J. Law Econ. Org. 11.
Teece, D.J., 1982. Towards an economic theory of the multiproduct firm. J. Econ. Behav. Organ. 3.
Walker, G., Weber, D., 1984. A transaction cost approach to make-or-buy decisions. Adm. Sci. Q. 29.
Wernerfelt, B., 1984. A resource-based view of the firm. Strateg. Manag. J. 5 (2), 171–180.
Williamson, O., 1971. The vertical integration of production: market failure considerations. Am. Econ. Rev. 61 (2), 112–123.
Williamson, O., 1975. Markets and Hierarchies, Analysis and Antitrust Implications : A Study in the Economics of Internal Organization. Free Press.
Williamson, O., 1985. The Economic Institutions of Capitalism: Firms, Markets. Free Press, Relational Contracting.