CAGE

Community Networks and Trade

CAGE working paper no. 649

November 2022

Johannes Böken,
Lucie Gadenne,
Tushar Nandi,
Marta Santamaria

WARWICK
THE UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

Community Networks and Trade

Johannes Böken, Lucie Gadenne, Tushar Nandi, Marta Santamaria*

November 2022

Abstract

Do community networks shape firm-to-firm trade in emerging economies? We study the role of communities in facilitating firm-to-firm trade and firm outcomes using data on firm-to-firm transactions and firm owners' community (castes) affiliations for the universe of medium- and large- sized firms in West Bengal, India. We find that firms are substantially more likely to trade, and trade more, with firms from their own caste. Studying the mechanisms underlying this effect, we find evidence consistent both with castes alleviating trade frictions and taste-based discrimination by firms against those outside their community. Guided by these stylized facts, we develop a model of firm-to-firm trade in which communities affect pair productivity and matching costs and estimate the model using our reduced-form estimates. A counterfactual extending the positive effects of castes on trade to *all* potential supplier-client pairs would increase the number of network links by 60% and increase average firm-to-firm sales by 20%.

*Böken and Santamaria: University of Warwick. Gadenne: Queen Mary University of London, Institute for Fiscal Studies and CEPR. Nandi: IISER Kolkata. We would like to thank Guilhem Cassan, Dan Keniston and Tatjana Kleineberg for sharing their caste and surname data, as well as David Atkin, Stefano Caria, Mirko Draca, Jonas Hjort, Clement Imbert, Rocco Macchiavello, Roland Rathelot and numerous seminar participants at the University of Warwick, Queen Mary, the Geneva Trade and Development Workshop (GTDW), University of Geneva and LSE Management lunch for helpful comments.

1 Introduction

Most firms in the developing world are small; this limits their capacity to innovate, export or even survive (Hsieh and Olken, 2014; Atkin and Khandelwal, 2020; Ciani et al., 2020). There is growing evidence that production networks matter for firm growth: the clients and suppliers firms trade with affect firm size (Bernard et al., 2022), productivity (Atkin et al., 2017; Antràs et al., 2017; Alfaro-Ureña et al., 2022) and access to credit (McMillan and Woodruff, 1999). Yet we know very little about the determinants of firms' production networks in developing countries.

One potential such determinant are communities. A large literature has shown that internal cooperation within communities supports domestic economic networks, and facilitates firms' access to credit (Fafchamps, 2000; Fisman et al., 2017), investors (Hjort, 2021), insurance (Mazzocco and Saini, 2012) and workers (Munshi and Rosenzweig, 2016; Caria and Labonne, 2022). Cross-country immigrant communities shape export flows across countries (Greif, 1993; Gould, 1994; Rauch, 2001), particularly in the context of high contractual frictions (Rauch and Trindade, 2003). Communities could similarly facilitate economic interactions in firm-to-firm networks in developing countries, where such frictions loom large.

This paper considers the role of community (caste) networks in India in shaping firm-to-firm trade. Using panel data on firm-to-firm transactions and information on the firm owner's community, we find that firms are substantially more likely to trade, and trade more, when they belong to the same community. We find that a large share of the effects of communities on trade is likely due to communities alleviating frictions, although we also find evidence consistent with taste-based discrimination. Motivated by these stylized facts, we build and estimate a model of network formation in which communities affect both the productivity of a client-supplier relationship and the cost of forming the relationship. This enables us to quantify the effect of extending the positive effects of castes on trade to all potential supplier-client relationships on production networks, trade and the aggregate economy.

Our contribution is threefold. First, we systematically document the effect of caste networks on firm-to-firm trade and provide evidence on the mechanisms underlying this effect. To do so, we use administrative data on the universe of firms paying taxes in the state of West Bengal, India, between 2010 and 2016, which enables us to overcome two key observational constraints. First, the data contains annual transactions between firms, enabling us to map firms' production networks. Second, we obtain the firm owners' names by searching firms in the official registries and assign each last name to a caste community (*jati*) using the anthropological literature (Singh, 1996). Jatis are the bedrock of India's social architecture and support economic and social networks (Munshi, 2019). Our data contains 106,775 firms in 723 castes, and over 200 million potential trade relationships.

We find large effect of castes on trade, both at the extensive and intensive margin. Firms within caste networks are likely to make similar industry and location choices, and experience correlated shocks, so we control flexibly for the joint locations and products sold by suppliers and clients, and allow for arbitrary shocks to all trading partners over time. We find that being in the same caste doubles the probability that two firms trade, and, when they do, increases trade volumes by close to 20%.

Why do castes affect firm-to-firm trade? Our second contribution is to shed some light on the mechanisms that underlie the effects of castes on trade. We first consider whether our estimates of caste effects on trade are larger for trading relationships where we expect frictions to be more severe. We find that castes matter more for the trade of products that are relationship-specific (for whom the hold-up problem is more of a concern), and when trading partners are located in areas with worse-performing courts (making the formal enforcement of contracts harder). Our results imply that at least two-third of the overall effects of castes we estimate can be explained by castes alleviating trading frictions. We then consider whether part of the effect of caste could be due to taste-based discrimination by firms against those outside their caste. We test for Becker (1957)'s argument that firms with strong discriminatory preferences will eventually be forced out of competitive markets, by looking at the survival rate of firms as a function of their within-caste preferences, in the spirit of Weber and Zulehner (2014). We find that firms with high preferences for trading with their caste, relative to their industry average, are indeed more likely to exit. Finally, we find no evidence of inaccurate statistical discrimination, as defined by Bohren et al. (2019b), in our context.

Our third contribution lies in the quantification of the effect of castes on the aggregate economy. To do this, we build on the quantitative production network model developed by Bernard et al. (2022) that features a continuum of firms with heterogeneous productivity and relationship capability, and endogenous match formation. We extend this framework to allow for a third source of heterogeneity coming from the firm's community (caste) affiliation. Castes shape production networks in two ways. First, we allow castes to affect the productivity of any supplier-client pair. Second, firms trading within their caste also face a caste-specific matching cost: castes may facilitate the formation of new relationships.

We estimate the model parameters using simulated method of moments. Whilst the estimation of all parameters is simultaneous, we leverage our empirical estimates on the effect of castes on trade to estimate the caste-specific model parameters. As castes facilitate trade in the model, we use the share of these estimates that can be attributed to castes alleviating frictions as (conservative) moments in our estimation. We find that castes increase match productivity and substantially decrease matching costs. Firm pairs in the same caste are 3% percent more profitable when trading and face matching costs that are 86% lower than firms in different castes. Importantly, we show that both caste parameters are necessary to match the data: a model allowing castes to only affect pair productivity, or only matching costs, is unable to replicate our reduced form results and our untargeted moments.

Finally, we use the estimated model to quantify the role of communities in shaping aggregating outcomes such as the number of connections within production networks and welfare. Our main counterfactual considers the effect of extending the positive effect of castes on trade to *all* potential supplier-client pairs. We find large aggregate effects. Welfare increases by 32%. This is due to a 63% increase in the number of network connections, and an input price reduction of 22%. The higher trade productivity and lower matching costs increases average firm-to-firm sales by 19%. These large effects are due to the fact that this counterfactual effectively reduces trading frictions for over 96% of potential supplier-client pairs , as same caste pairs represent only 3.6% of all potential pairs in our data.

This paper contributes to the well-established literature considering how firms in developing economies respond to contractual frictions by establishing relational contracts (see Macchiavello, 2022, for a review), vertically integrating (Woodruff, 2002; Boehm and Oberfield, 2020; Hansman et al., 2020) or building market reputation (Banerjee and Duflo, 2000; Macchiavello and Morjaria, 2015). In particular, a growing literature studies how firms leverage community networks to mitigate these frictions (see for example Fafchamps, 2000; Fisman, 2003; Banerjee and Munshi, 2004; Munshi, 2011, 2014; Fisman et al., 2017; Dai et al., 2020). Our paper builds on this literature by providing the first estimates of the role of community networks in shaping firm-to-firm trade, and quantifying the effects of these networks on the aggregate economy. When investigating mechanisms we find that a large share of the effect of caste on trade that we observe can be explained by castes alleviating market frictions, in line with this literature.

We also find evidence consistent with taste-based discrimination along caste lines, suggesting that community networks may be simultaneously both facilitating trading relationships and distorting trade away from firms with small networks. This result speaks to the literature documenting the negative effects of ethnic heterogeneity and growth (Alesina and Ferrara, 2005) by providing micro-evidence on how communities can erect new trade barriers whist alleviating trading frictions. Our results are particularly related to Hjort (2014) who finds that workers in Kenyan firms lower their own pay to decrease that of their non-coethnic colleagues. Our evidence that firms with stronger own-caste preferences are more likely to exit similarly point to agents' willingness to sacrifice some economic gains to sustain their community networks (see also La Ferrara, 2002).

Our results also build on the existing literature on the role of communities in shaping trade, which typically focuses on how immigration networks and cultural proximity across countries affect international trade (Greif, 1993, 2006; Rauch, 2001; Rauch and Trindade, 2003; Bandyopadhyay et al., 2008; Guiso et al., 2009)[1]. Our focus on intra-national trade and data on firm-to-firm transactions enables us to provide micro-level evidence of the size of community effects on trade, as well as provide evidence regarding the mechanisms underpining these effects.

Finally, we contribute to the literature exploring the aggregate implications of

[1] An exception is (Combes et al., 2005) who find that migrant communities and business networks in France increase regional trade flows within the country

production networks, and their role in shaping firm outcomes such as firm size (Bernard et al. (2022)) and firm productivity (through sourcing decisions (Antràs et al., 2017), information flows from international buyers (Atkin et al., 2017),) and interactions with multinational corporations (Alfaro-Ureña et al., 2022)). Previous studies have shown that network formation and development is subject to high costs to add and replace suppliers (Bernard et al. (2019) in Japan, Huneeus (2018) in Chile, Startz (2021) in Nigeria). These adjustment costs may be higher in the context of low and middle income countries (Atkin and Khandelwal (2020)), where contract enforcement institutions may be weak (Nunn, 2007) or congested (Boehm and Oberfield, 2020). We extend the parsimonious network formation model in Bernard et al. (2022) to model the role of communities in facilitating trade in the presence of contractual frictions.[2] We find that communities in India (castes), are quantitatively important in reducing frictions associated with relationship-specific inputs and contract enforcement and we show that reducing these frictions for all firm-pairs would have large aggregate implications for firm-to-firm trade and connections, as well as for welfare.

The rest of the paper is organized as follows. Section 2 describes our context of study, highlighting the role of castes in shaping economic outcomes in India, and the data we use. Section 3 presents new stylized facts on the effect of community networks on trade and evidence regarding the mechanisms underlying these effects. Section 4 presents our model, section 5 our model estimation strategy, results and counterfactual exercises.

2 Context and data

Our context of study is West Bengal, a large Indian state with 90 million inhabitants and a GDP per capita of USD 8,200 (ppp) in 2020, similar to India's national average. Our period of study is 2010-2016.

[2]Recent theoretical advances in this literature have introduced two-sided heterogeneity, in firm productivity and matching costs, to be able to replicate the main patterns in firm-to-firm networks in Norway (Bernard et al. (2018)) and Belgium (Bernard et al. (2022)).

2.1 Community networks in India

India's social architecture is organized around thousand castes or *jatis*. Internal cooperation within castes supports economic networks: marriages are typically within castes, informal loans and insurance mechanisms are concentrated within castes and castes historically determined individuals' occupation and location choices (see Munshi, 2019, for a review of the role of caste in Indian society). Whilst the concept of caste originates in Hinduism, it has extended across other religions, with non-Hindu castes playing a similar role as Hindu castes in Indian society today (Cassan, 2020).

There is evidence that caste networks help alleviate market frictions in credit markets (Fisman et al., 2017), labor markets (Munshi and Rosenzweig, 2016) and insurance markets (Mazzocco and Saini, 2012): cultural proximity between caste members reduces asymmetric information, allowing transactions to occur in contexts with severe informational and contractual frictions and thereby increasing market efficiency. The existence of caste networks could however simultaneously lead to individuals transacting more within caste for preference-based reasons, leading to discrimination and ultimately resource misallocation as individuals' economic opportunities are constrained if they do not belong to the 'right' caste. Caste networks have been shown to lead to such inefficiencies in capital markets (Banerjee and Munshi, 2004), groundwater trade (Anderson, 2011) and education decisions (Munshi and Rosenzweig, 2006).

2.2 Data on production networks

We consider how caste networks shape firm-to-firm trade by using detailed data on firm-to-firm transactions matched with information on firm owners' castes. We use administrative data on firm-level tax returns and tax registration information obtained from the West Bengal Directorate for Commercial Taxes for the fiscal years 2010-2011 to 2015-2016, containing information on the universe of all firms paying Valued-Added-Taxes (VAT) to the state over the period.

The tax returns data documents all transactions between firms paying VAT in West Bengal: both firms involved in the transaction report the annual transaction amount as well as the tax identification number of their client or supplier. The tax registration data contains information on firms' locations (1088 unique postcodes) and the products sold by the firms which we classify using India's National Industry Classification (NIC) into 162 product codes. For 77% of firms in our data the product codes are available at the detailed 4-digit level, for the remainder we use 3-digit codes. Controlling for detailed product information affects the interpretation of our estimates of caste effects on trade so we present robustness checks using only firms for which detailed product level information is available below. This data is described in more detail in Gadenne et al. (2022).

2.3 Other data

We use several other datasets to consider whether castes play a different role for trading relationships facing more severe contractual frictions. To proxy for the difficulty of enforcing contracts legally, we construct a measure of local court congestion. We use data on 2.6 million cases from District and Session courts in West Bengal between 2010 and 2018, collected from the Indian e-courts platform by Ash et al. (2021). Each case record includes information on the court's district, the filing date and, if applicable, the decision date. Our preferred measure of court congestion is an indicator of whether a case had been decided two years after being filed, as our period of study ends in 2016. We aggregate these into an ex-post probability that a case filed in a given district and fiscal year will be decided within the next two years.[3]

We also consider whether castes affect trade in relationship-specific products differently. We use the classification in Rauch (1999) to characterize all products as either homogeneous (traded on an organized exchange or with a reference price) or relationship specific. For our measure of goods' relationship-specificity, we use the concordance tables from Liao et al. (2020) to obtain the share of relationship-specific inputs within each NIC 4-digit code.[4]

2.4 Variable and sample creation

Our main variable of interest is an indicator for whether two firm owners belong to the same caste community. Firms' tax identification number is public knowledge in India, information on the firm owner's name can be obtained by querying

[3]We consider court congestion in the client's district, in 57% of cases the supplier and the client are located in the same district.

[4]We concord from the original SITC Rev 2 codes to NIC via 6-digit NAICS codes

the firm ID on a public database. We follow Cassan et al. (2021) in using the systematic classification of Indian last names into various tribes and communities, including 2,205 castes (or 'main communities') developed by the People of India project in their 1985 Anthropological Survey of India (Singh, 1996). The merge between last names and castes is sometimes not unique: 65% of firm owners' last names in our data are associated with more than one caste. When this occurs, we allocate the first caste in alphabetical order to each last name. We discuss the robustness of our results to alternative methods of allocating last names to a caste below. Using this method, we obtain information on the firm owner's caste for 75% of firms in the administrative tax data, see Appendix A for more details.

Our final firm-level sample consists of 106,775 firms allocated to one of 723 unique caste communities. The average community size is 148 firms, Appendix Figure A2 plots the distribution of the number of firms per caste. When considering the effect of caste on the intensive margin of trade below, we consider the sample of all 1,461,018 transactions recorded between these firms over our six year period, these transactions take place within 764,767 unique supplier-client pairs.

Given the size of our data, the universe of all potential supplier-client matches is extremely large (22 billion potential pairs) and computationally intractable. To consider the effect of castes on the extensive margin of trade we therefore define a 'potential trade' sample. We first define a client (supplier) as a firm observed at least once on the purchasing (selling) side of a transaction at least once in a given fiscal year - firms can be, and often are, both clients and suppliers. We then restrict the set of available suppliers for each client using information on the products sold by firms. For each client in our data observed trading with suppliers selling products P, we consider the set of all suppliers selling products P and randomly include 25% of them as 'potential suppliers' for this client. This 'potential trade' sample contains 202 million potential annual supplier-client pairs. We consider instead the full set of 'potential suppliers' for a single year as a robustness check. This sample definition essentially assumes that the set of products that firms trade is fixed, and not affected by caste networks. To relax this assumption and allow, for example, for the possibility that firms adjust their input mix based on the caste of their potential suppliers, we consider a second, large potential trade sample based on 'recipes' in the spirit of Atalay et al. (2019) as a robustness check. To construct this sample, we consider as potential suppliers for each client selling products P' a random 25% of the set of all suppliers seen trading with any client selling products P'[5].

Table 1 presents descriptive statistics for our transaction data. In the first panel we see that firms have an average of 3 suppliers and 3 clients, but nearly 400-500 potential suppliers and clients to choose from, of which 15-18 are from their own caste. There is substantial entry and exit in our data - we observe firms for an average of 3 to 4 years. The potential trade data, described in the last panel, shows that the trade matrix is very sparse: only 0.7% of the potential supplier-client pairs in our data are observed trading, and 3.6% of potential supplier-client pairs are from the same caste.

3 Empirical evidence on community networks and trade

3.1 Stylized facts

Graphical evidence. Figure 1 presents graphical evidence on the role of castes in firm-to-firm trade using our sample of potential trade. We plot the relationship between how much a firm is observed trading with others in the same caste and how much it could potentially trade with same-caste firms conditional on the distribution of castes in the potential trade data. Panel a) plots the share of firms' input purchases from same-caste suppliers as a function of their potential same-caste input share: the share of same-caste suppliers in all their potential suppliers. Panel b) similarly plots the share of firms' sales going to same-caste clients as a function of their potential same-caste sales share. Potential clients and suppliers are as defined above, and weighted by their average network sales, so that these potential input and sale same-caste shares can be interpreted as how much firms would trade within their caste if they randomly chose their trading partners (Bernard et al., 2019, use a similar approach to consider how distance affects firm-to-firm transactions).[6]

[5] For tractability reasons, we only consider recipes that are used by at least 1% of an industry's suppliers or clients, where each firm is weighted by its total sales.

[6] Our baseline potential trade sample allocates potential suppliers to each client and uses the subsample stratified by client; it is used to produce panel a). To produce panel b) we instead use a subsample stratified by supplier.

We see that firms systematically trade a lot more within their caste than they would if trading relationships were randomly chosen: each point on both panels is clearly above the 45 degree line. This is true both for firms in large castes (those with high potential same-caste shares) and for firms in smaller castes. On average, firms' potential same-caste input share (sales share) is 4.5% (5.1%) but the average observed same-caste input and sales shares are more than twice as large, at 10.6% and 12.4%, respectively.

This graphical evidence is a first indication that castes affect firm-to-firm trade, but it could be confounded by firms in the same caste making similar location or product choices. In what follows we turn to a regression framework to quantify the effect of caste on both the intensive and extensive margin of trade whilst controlling flexibly for all determinants of trade that could be correlated within caste networks.

Regression evidence. To measure the effect of caste networks on trade, we estimate the following gravity equation augmented to allow for destination (client i) and origin (supplier j) fixed effects that vary across years t:

$$ln(Y_{ijt}) = \beta \mathbb{1}(c_i = c_j) + \gamma X_{ijt} + \mu_{it} + \mu_{jt} + \epsilon_{ijt} \tag{1}$$

where Y_{ijt} is log sales from j to i in year t when we consider the intensive margin of trade, and an indicator equal to 1 if firms are trading when we consider the extensive margin, $\mathbb{1}(c_i = c_j)$ is an indicator equal to 1 if the owners of firms i and j belong to the same caste, μ_{it} and μ_{jt} are, respectively, supplier and client fixed effects interacted with year fixed effects, and X_{ijt} is a set of controls defined at the ij pair and period t level, discussed below. We use the potential trade sample defined above when considering the extensive margin, and the sample of all positive sales when considering the intensive margin. Standard errors are two-way clustered at the level of the client and the supplier.

Castes play an essential role in economic and social interactions so we could see firms trading more with others in their own caste (a positive β) even in the absence of caste playing a direct role in firm-to-firm trade for. First, castes are known to affect occupational choice (Cassan et al., 2021) so firm owners of the same caste may find themselves in the same supply chains because of their sector choice. To control for this channel we include fixed effects for the interaction of both firms' products. Second, some castes are concentrated geographically, so firm owners may choose to trade more with others in their caste simply because of lower transport costs. We account for this channel by controlling non-parametrically for the effect of distance on trade by including fixed effects for the interaction of the location (postcode) in which both firms are located (Head and Mayer, 2014). Third, firm owners in the same caste may face similar aggregate shocks: the role of castes as providers of credit and insurance implies that unobserved shocks to firm owners in the same caste are likely to be correlated. Allowing for arbitrary shocks over time to both clients and suppliers ensures that our estimates of β cannot be driven by such caste-level shocks.

Table 2 presents our results on the effect of castes on the extensive margin of trade, obtained by running specification (1) on the potential trade sample. Coefficients are rescaled by the average probability that two firms trade in our potential trade sample, so they can be interpreted as the effect of caste on the probability that two firms choose to trade. We see that being in the same caste increases the probability that two firms trade by 130% in the specification with no controls (column 1). Adding interactions for the location of the client and the supplier (in column 2), firm$\times$ year fixed effects (in column 3) and interactions for the products sold by the client and the supplier (in column 4) all decrease the effect of two firm owners being of the same caste on the probability that they trade, as expected. Our preferred estimate, in column (4) indicates that being of the same caste doubles the probability that two firms trade.

Estimates of the effect of castes on the intensive margin of trade in Table 3 are obtained by running specification (1) on the sample of all positive trades. We see again that controlling for firms' joint locations, products sold, and (in particular) allowing for arbitrary shocks to clients or suppliers decreases the effect of castes on trade, but our preferred estimate in column 4 indicates that firm owners in the same caste trade substantially more (18%) with each other. Overall, results in Tables 2 and 3 indicate that castes substantially increase both the probability that two firms trade and, when they do, how much they trade, even when controlling non-parametrically for other firm choices (location, products) that are likely correlated within castes and allowing for arbitrary caste-level shocks. These effects of castes on trade are very large, though of comparable magnitude to the estimated effect of castes on another 'matching market': Banerjee et al. (2013) find that the probability that individuals respond to letters received in response to matrimonial ads by 45-60% when the letter-writer is of the same caste.

Tables A3 and A4 consider the robustness of our estimates of the effect of castes on trade to our data construction, sample and specification choices. In columns 1 to 7 we exclude firms whose products are not defined at the detailed 4-digit level, consider several alternative assignments of surnames to castes, exclude the largest 3 castes and all firms with no same-caste potential partners, and consider standard errors clustered two-way at the level of the client's industry and location. Results are remarkably similar across specifications and samples. For the extensive margin results we additionally consider alternative ways to define firms' sets potential trading partners, looking at one year (2013) only to keep manageable sample sizes. We find that keeping all potential trading partners (instead of a random 25% sample) does not affect our results. When using the 'recipes' potential trading partners definition in the spirit of Atalay et al. (2019), described above, we no longer constrain a firm's potential suppliers to sell products this firm is observed buying. Our estimate of the effect of castes in this sample could therefore also captures the fact that firms' caste networks may also affect which inputs they buy. We find a larger estimate using this sample than with our baseline sample, suggesting this may be the case. Finally, Table A6 tests whether the caste effects we observe reflect the role of *jati* networks or simply a preference for trading within the larger caste *varna* groups (there are four varna groups, with an additional *dalit* group so we allocate each firm to one of five large groups). We find a small effect of *varnas* on trade of roughly 7% the magnitude of our baseline caste effect, which remains unchanged when allow *varnas* to also affect trade.

3.2 Mechanisms

Having established a large effect of caste on the extensive and intensive margins of firm-to-firm trade, we now turn to discussing potential mechanisms which could explain why firms trade more within than across castes.

Contractual frictions. A large literature on client-supplier relationships in developing countries argues that contractual frictions loom large in this context (see Macchiavello, 2022, for a review), and that social networks can help alleviate these frictions by providing informal information and sanction mechanisms

(Greif, 1993, 2006). Given their importance in the organization of Indian society, caste networks could enable relational contracts to emerge, allowing for trading relationships to be sustained even in India's notoriously weak contract enforcement environment (Boehm and Oberfield, 2020). We test this hypothesis in two ways. We first consider whether the effect of caste on trade varies for inputs that are more relationship-specific, using the classification from Rauch (1999) to attribute a relationship-specificity score to the products sold by the supplier, as explained above. Hold-up problems are more likely to arise with relationship-specific goods (Iyer and Schoar, 2015); if castes networks sustain informal enforcement mechanisms, we expect castes to increase trade in these products more than trade in homogeneous products. Second, we follow Boehm and Oberfield (2020) and use court congestion at the district level to proxy for the strength of formal enforcement mechanisms. If castes help enforce contracts, the caste effect on trade should be higher in areas in which formal enforcement channels are weaker.

Table 4 shows evidence consistent with caste networks alleviating contractual frictions. We see that the effect of caste on trade is higher in areas with more congested courts and for relationship specific products, for both the intensive and extensive margins of trade. A one standard-deviation increase in court congestion (in the traded product's relationship specificity score) increases the intensive margin caste effect by 1.4 (2.4) percentage points and the extensive margin effect by 10.4 (9.6) percentage points. The coefficients for the same caste indicator in the last two columns can be interpreted as the effect of castes for trade that occurs in contexts with low (or no) contractual frictions: trade of perfectly homogeneous products in areas in which courts complete all cases within two years of them being filed. We see that the effect of castes on such trade, whilst much lower than the average effect, is still economically and statistically significant at roughly one-third of the average effect on both the extensive and intensive margins.

Taste-based discrimination. We test for the existence of taste-based discrimination by looking at the effect of a firm's caste preference on its survival probability in later years.[7] This test is inspired by Weber and Zulehner (2014) who build on the argument in Becker (1957) that firms with strong discriminatory preferences will forego profits by submitting to these preferences and, in competitive markets, will eventually be forced out. In our context, this implies that firms with strong own-caste preferences (proxied for by an own-caste input share above industry average) will be less likely to survive. Specifically, we estimate a Cox regression relating firm exit hazard to same-caste input share, relative to the industry average, and additional controls age and size. We furthermore control flexibly for industry, location, and size using stratification, only comparing firms within groups of similar firms based on these characteristics, therefore allowing the baseline hazard to vary non-parametrically between strata. In our preferred specification, column (4) of Table 5, we only compare firms from the same post-code, selling the same product and from the same decile of the size distribution, yielding a total of 59,936 unique groups.

Results on firm survival as a function of their caste preference are presented in Table 5. We see that firms with strong preferences for trading within their caste are more likely to exit. A one standard deviation increase in a firm's own-caste input share (relative to the industry average) is associated with a 2.1% higher risk of exit in every time period. This evidence is consistent with part of the observed effect of castes on trade being explained by firms being prejudiced against trading outside of their caste networks. We find a similar relationship for the relationship between firm exit hazard and own-caste sales share in Appendix Table A2 (5.7% higher risk of exit for a 1 standard deviation increase in own-caste sales share).

Inaccurate statistical discrimination. Finally, we note that statistical discrimination, whereby firms are reluctant to trade with others belonging to a group with worse outcomes in expectation, is unlikely to explain the patterns we observe. This is because our caste effects reflect symmetric preferences for in-group interaction, not asymmetric preferences with most firms preferring to avoid a specific

[7]Firms that exit our data may not necessarily stop operating - they may merely have chosen to stop paying VAT to the state government. Whilst informal firms are common in India, it is relatively easy for the tax authorities to find firms, and make them pay their taxes, if they were registered with the tax authorities in the previous year. Firms below a certain size do not have to pay VAT, and while they are still required to file taxes, not doing so may be tolerated by the authorities. Overall, an exit from our data can be interpreted as either a sign that the firm stops operating, or that it becomes so small it thinks the tax authorities will ignore the fact that it has stopped filing taxes. In both cases, an exit is a sign of poor profitability.

group with worse outcomes. *Inaccurate* statistical discrimination, as defined by Bohren et al. (2019a), could however play a role: firms could hold biased beliefs about those outside their caste. They could incorrectly expect them to be worse-performing trading partners by, for example, providing worse-quality goods compared to trading partners from their own caste. Bohren et al. (2019b) show that looking at the dynamic patterns of discrimination helps uncover biased beliefs, because individuals holding biased beliefs about others' potential performance will change these beliefs as more performance is observed. Building on their intuition, we consider how the caste effect on trade changes over time, as firms in a trade relationship learn about each other's suitability as trading partners. If firms systematically and incorrectly believe that partners from their own caste are better performers (relative to partners from other castes) than they truly are, we should see within-caste relationships fail to thrive relative to across caste relationships. This would in turn lead to the effects of caste on trade in Tables 2 and 3 to diminish in pairs that have been trading longer.

Figure 2 uses the panel dimension of our data to investigate whether the effect of caste on trade fades over time. Looking at a sample of newly formed pairs we find that this effect is remarkably stable over the 5 years we can observed them for. In panel a) we see that, if anything, same-caste pairs are slightly more likely to survive - the opposite of what we would expect if firms were de-biasing their incorrect beliefs over time. The intensive margin effect of caste on trade in panel b) is fairly stable over time, with firms buying 12-25% more from same-caste suppliers regardless of how long the trading relationship has been in existence. Overall, we see no evidence of the effect of caste on trade decreasing as firms learn more about each other. This suggests inaccurate statistical discrimination isn't driving our results. Consistent with this, in Table A5 we find no evidence that caste networks affect trade less when either the client or the supplier has more experience or a better-established reputation (proxied for by time since registration with the tax authorities).

Overall, our results indicate large effects of caste networks on trade at the intensive and extensive margin. We find evidence consistent both with castes alleviating market frictions, and with part of the caste effect being driven by taste-based discrimination. Our next section builds a model of firm-to-firm transactions to enable us to quantify the overall effect of castes on trade.

4 A model of firm-to-firm networks with communities

We use a theoretical framework of supplier-client networks with two-sided firm heterogeneity and endogenous match formation, built on Bernard et al. (2022). Our application of the framework to the context of West Bengal, India, is motivated by three empirical patterns in our firm-to-firm data.

Fact 1: The distributions of firm sales and supplier-client links are highly dispersed. The distributions of firm size (firm sales), number of suppliers and number of clients have a high dispersion, spanning several orders of magnitude (see Figure A4 in the Appendix). The largest firms sell thousands of times more than the average firm and have thirty to fifty times more suppliers or clients. This high dispersion in the distribution of sales and links suggests that firm heterogeneity is high in West Bengal, calling for a model which accounts for firm-heterogeneity.

Fact 2: Firms with more customers have higher sales and higher sales per customer. We find that large firms have more clients and also higher sales per client (see Figure A5 in the Appendix). This pattern suggests that firm size affects both the ability to adopt more clients, and the ability to sell more to those clients.[8]

Fact 3: Suppliers with more customers match with customers who have fewer suppliers on average. Our data features negative degree assortativity: Firms with fewer customers match with well connected customers, while firms with many customers match with less-well connected firms on average (see Figure A6 in the Appendix). This is a well-known feature of business-to-business networks which motivates the choice of a parsimonious model of firm-matching where a match is formed when the profits of the match are larger than the cost of the match.

In the remaining part of this section, we develop a firm-to-firm model that is consistent with these data patterns and will also feature firm community-affiliation.

[8]The positive correlation between size and the sales per client suggests a strong connection between firm-heterogeneity, firm size and firm connections. Bernard et al. (2022) find a negative correlation between network sales per customer and number of customers, suggesting that firms with many customers are not able to sell more to each customer, which implies a negative correlation between productivity and matching ability. Our data pattern, on the contrary, suggest a lack of correlation between matching ability and productivity, similar to patterns uncovered in Norwegian data in Bernard et al. (2018).

4.1 Theoretical framework

We first present the model conditional on a fixed firm network, and subsequently introduce a parsimonious firm-to-firm matching model.

There are three sources of firm heterogeneity. First, firms have different productivity levels, that help them produce inputs more efficiently. Second, firms have different relationship capabilities, that allow them to create new firm-to-firm relationships by paying different fixed costs. Finally, firms belong to a specific caste community. We specify how castes affect trade in the following subsection.

4.1.1 Technology and Demand

The economy is formed by a unit continuum of firms, each with the following production function:

$$y(i) = \kappa z(i) l(i)^{\alpha} v(i)^{1-\alpha}, \tag{2}$$

where $y(i)$ is the quantity of output produced by firm i, $z(i)$ is the productivity, $l(i)$ is the amount of labor used by firm i, α is the labor share, and $\kappa > 0$ is a normalization constant.[9] $v(i)$ is the bundle of intermediate inputs used by the firm in production, given by:

$$v(i) = \left(\int_{\mathcal{S}(i)} \nu(i,k)^{\frac{(\sigma-1)}{\sigma}} dk \right)^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where $\nu(i,k)$ is the quantity that firm i purchases from firm k, $\mathcal{S}(i)$ is the set of suppliers available to firm i, and $\sigma > 1$ is the elasticity of substitution across suppliers within the sector. The input price index associated with the CES input demand function is given by $P(i) = \left(\int_{\mathcal{S}(i)} p(k)^{1-\sigma} dk \right)^{1/(1-\sigma)}$, where $p(k)$ is the price charged by supplier k. This input price index will be lower when firm i is able to match with efficient suppliers (that charge a low price $p(k)$ for their inputs) and when firm i can source from a larger set of suppliers (thanks to the properties of the CES input bundle aggregator).

Following Bernard et al. (2022), we choose the wage as the numeraire ($w = 1$) and we can express the marginal unit cost at which firm i can sell a unit to firm j

[9] The constant is $\kappa = \alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}$

as:

$$u(i,j) = \frac{P(i)^{1-\alpha}}{z(i)\delta_z^{1-\mathcal{C}_{i,j}}}, \tag{4}$$

where $\mathcal{C}_{i,j} = 0$ if the firms belong to the same caste, 1 otherwise. Parameter $\delta_z^{1-\mathcal{C}_{i,j}}$ is the productivity effect of same-caste trading. We interpret this pair-productivity term as an effect coming from the different trust or risk associated with same-caste trading. In contexts with high frictions and low enforcement, clients may want to pay slightly higher prices to ensure that the supplier fulfills the contract, especially when interactions are repeated (see Startz (2021)). If same-caste pairs have higher trust or lower risk, the contracting premium will be lower (which will translate into a higher pair-productivity effect). We will estimate $\delta_z^{1-\mathcal{C}_{i,j}}$ to explore this hypothesis. As we can see, both the firm's productivity and the firm's supplier connections help reduce the marginal cost of production.

Final demand Final consumers have a CES utility function with the same elasticity of substitution σ across output varieties. We assume that the representative consumer is the shareholder of all firms, so that aggregate profits Π become part of consumer income. Aggregate income X is therefore the sum of aggregate labor income and aggregate corporate profits, $X = wL + \Pi$, where L is inelastically supplied labor.

4.1.2 Firm-to-Firm Sales

Each firm faces demand from other firms, as well as from final consumers.[10] Given our assumption about the production function and the demand for intermediates, we can solve for the sales from firm i to firm j:

$$m(i,j) = p(i,j)^{1-\sigma} P(j)^{\sigma-1} M(j), \tag{5}$$

where $M(j)$ are total intermediate purchases by firm j, $\int_{S(j)} m(k,j)dk$ and $p(i,j)$ is the price at which firm i sells one unit to firm j.

The market structure of monopolistic competition means that firms will choose to charge a constant mark-up over marginal costs, $p(i,j) = \mu \times u(i,j)$, where

[10]Each firm sells its final output as intermediate input to other firms and to final consumers. The sourcing and producing of intermediate varieties happen simultaneously.

$\mu = \sigma/(\sigma-1)$. After rearranging, sales from i to j can be expressed as:

$$m(i,j) = \left[\frac{z(i)\delta_z^{1-\mathcal{C}_{i,j}}}{\mu P(i)^{1-\alpha}} P(j)\right]^{\sigma-1} M(j) \tag{6}$$

Taking logs of equation (6), it is easy to see that the model delivers a similar log-linear expression of firm-to-firm sales as the one we used in our empirical section, equation (1). The caste dummy will identify the pair-productivity effect of same-caste trading.

4.1.3 Equilibrium Conditional on Network

The equilibrium can be computed in two separable steps. First, we describe the equilibrium given a fixed firm-to-firm network. Then, we solve for the equilibrium network introducing endogenous match formation.

A firm i is characterized by the tuple $\lambda = (z; F)$, where z is productivity and F is a relationship fixed cost, paid in units of labor. In the model, z and F are potentially correlated, and dG (λ) denotes the (multivariate) density of λ. We define the link function $l(\lambda, \lambda')$ as the share of supplier-client pairs (λ, λ') that match in a trade relationship.

Backward fixed point. For a given network structure, the equilibrium can be found by solving for two fixed points sequentially. Using the pricing rule $p(\lambda, \lambda') = \mu c(\lambda, \lambda')$ and the equation for the marginal unit cost (4), the input price index can be solved for by iterating on a backward fixed point problem:

$$P(\lambda)^{1-\sigma} = \mu^{1-\sigma} \int P(\lambda')^{(1-\sigma)(1-\alpha)} \left(z(\lambda')\delta_z^{1-\mathcal{C}(\lambda,\lambda')}\right)^{\sigma-1} l(\lambda', \lambda) dG(\lambda') \tag{7}$$

The input cost index of firm λ, $P(\lambda)$, depends on the input cost index and productivity of all its suppliers, $P(\lambda')$ and $z(\lambda')$.

Forward fixed point. Sales of a type-λ firm are the sum of sales to final and intermediate demand: $S(\lambda) = \mathcal{F}(\lambda) + \int m(\lambda, \lambda') l(\lambda, \lambda') dG(\lambda)$ where $m(\lambda, \lambda')$ now denotes sales by supplier λ to client λ'. Final demand is $\mathcal{F}(\lambda)$ = p $(\lambda)^{1-\sigma}\mathcal{P}^{\sigma-1}X$, with the consumer price index equal to $\mathcal{P}^{1-\sigma} = \int p(\lambda)^{1-\sigma} dG(\lambda) =$

$\mu^{1-\sigma} \int P(\lambda)^{(1-\sigma)(1-\alpha)} z(\lambda)^{\sigma-1} dG(\lambda)$. Also note that total input purchases are $M(\lambda) = S(\lambda)(1-\alpha)/\mu$. Using this equation together with the expression for the marginal unit cost (4) yields:

$$S(\lambda) = \left(\frac{\mu z(\lambda)}{P(\lambda)^{1-\alpha}}\right)^{\sigma-1} \left(\frac{X}{\mathcal{P}^{1-\sigma}} + \frac{1-\alpha}{\mu} \int \frac{S(\lambda')}{P(\lambda')^{1-\sigma}} \delta_z^{1-\mathcal{C}(\lambda,\lambda')(\sigma-1)} l(\lambda,\lambda') dG(\lambda')\right) \tag{8}$$

Sales of a type-λ firm depend on final demand, X, the productivity and input price index of the firm itself, $z(\lambda)$ and $P(\lambda)$, and the sales and input prices of its customers, $S(\lambda')$ and $P(\lambda')$. In addition, the sales of each supplier will be affected by whether the supplier and the client have the same caste affiliation. The equilibrium exists and is unique (Bernard et al., 2022).

4.1.4 Firm-to-Firm Matching

We now consider the general equilibrium when the firm-to-firm network is endogenous and suppliers match with clients if and only if the profits from doing so are positive. The supplier incurs a relationship fixed cost $F\varepsilon$ for every client it chooses to sell to, where F varies across suppliers, and ε is an idiosyncratic shock that varies across supplier-client pairs.[11] In addition, the cost of creating a link differs when both firms belong to the same caste, so that the cost of creating a trade relationship is $F\varepsilon e^{\delta_F(1-\mathcal{C}_{i,j})}$, where $\mathcal{C}_{i,j} = 0$ if $caste_i = caste_j$, 1 otherwise. Creating a new trading relationship may be less costly with same-caste firms due to better information about the available trade partners or higher trust.

Given these assumptions on the matching technology, the share of supplier-client pairs (λ, λ') that match and trade with each other is:

$$l(\lambda,\lambda') = \int \mathbb{1}\left[ln\varepsilon < ln\pi(\lambda,\lambda') - lnF + (1-\mathcal{C}_{i,j})\delta_F\right] dH(\varepsilon), \tag{9}$$

where $\mathbb{1}[]$ is the indicator function, $dH(\varepsilon)$ denotes the density of ε, and the gross

[11] The introduction of this pair-specific shock is needed to smooth the problem and ensure that the matching function is continuous in the parameters of the model (Bernard et al., 2022)).

profits from the potential match are:

$$\pi(\lambda, \lambda') = \frac{m(\lambda, \lambda')}{\sigma}. \tag{10}$$

Notice that this expression will give us the trade probability for a pair type (λ, λ'). This trade probability depends on the profitability of the match and on the matching costs. It is worth highlighting that both caste parameters enter this equation and affect trade probabilities. The pair-productivity term δ_z, increases the match gross profits. The caste matching cost δ_F affects the matching costs. This expression shows why both caste effects could be driving the observed higher same-caste trade probability.

Given the gross profits, the matching costs and the pair-shocks, this link function is also a fixed point problem.

We can now explain how to solve for the general equilibrium, that nests the three fixed point problems. The algorithm is proposed by Bernard et al. (2022):

(i) Start with a guess for the link matrix.
(ii) Use equations (7) and (8) to solve for for $P(\lambda)$ and $S(\lambda)$ sequentially.
(iii) Calculate gross profits for all potential matches and compute the share of supplier-client pairs that match according to equation (9).
(iv) Go back to step (ii) and repeat until the link matrix converges.

4.1.5 Predictions on communities and trade in the model

The model delivers both an intensive margin equation (6) and an extensive margin equation (9) that we can compare to our reduced-form evidence.

1. Firm-to-firm sales

$$ln\Big(m(\lambda, \lambda')\Big) = \underbrace{(1 - \mathcal{C}(\lambda, \lambda'))}_{\text{Caste dummy}}(\sigma - 1)ln(\delta_z) + \underbrace{(\sigma - 1)ln\left[\frac{z(\lambda)}{\mu P(\lambda)^{1-\alpha}}\right]}_{\text{Client FE}} + \underbrace{ln\left[\frac{M(\lambda')}{P(\lambda')^{1-\sigma}}\right]}_{\text{Supplier FE}} \tag{11}$$

Conditional on client and supplier fixed effects, a positive coefficient on the caste-dummy is evidence of positive pair-productivity effect of same-caste trade, $\delta_z > 1$.

2. Firm-to-firm trading probabilities

$$l(\lambda,\lambda') = \int \mathbb{1}\left[ln\varepsilon < \underbrace{ln\left(\frac{m(\lambda,\lambda')}{\sigma}\right)}_{\text{Match profits}} - \underbrace{lnF}_{\text{Firm Matching cost}} + \underbrace{(1-\mathcal{C}(\lambda,\lambda'))\delta_F}_{\text{Caste matching cost}} \right] dH(\varepsilon), \tag{12}$$

A higher trading probability within-caste could be evidence of both a pair-productivity effect from same-caste trading ($\delta_z > 1$) and a lower matching cost for same-caste pairs ($\delta_F < 0$)

In the next section we explain how we estimate the model and use it for counterfactual exercises.

5 Estimation and Results

This section provides a model-based quantification of the aggregate effects of removing all inter-caste trading frictions. We start by estimating the model using our production network data and the estimated effects of castes on trade presented in section 3. This yields estimates of the underlying parameters governing the effect of castes on trade, δ_z and δ_F. We then use the model to consider two counterfactual scenarios.

5.1 Simulated Method of Moments

We estimate the model using simulated method of moments (SMM). We assume that firm productivity z and relationship capability F follow a log-normal distribution with expectations $\mu_{lnz} = 0$ and μ_{lnF}, standard deviations σ_{lnz} and σ_{lnF}.[12] We calibrate several parameters by drawing on the existing literature. We follow Boehm and Oberfield (2020) in calibrating the labor cost share α to the Indian context, De Loecker et al. (2016) to assign a value of the markup μ, Bernard et al. (2022) for the standard deviation of the idiosyncratic matching cost $\sigma_{ln\epsilon}$, and normalize aggregate final demand X to 1. We calibrate the distribution of firms' caste affiliations to replicate the caste distribution in our data, including the share of same-caste pairs in the potential trade data (3.57%, see Table 1). We simulate

[12] Unlike Bernard et al. (2022) we assume that z and F are uncorrelated, allowing for a correlation between these two dimensions of heterogeneity does not improve model fit.

the model with 300 firms, allocating the following caste communities: two large castes with 10% of firms each, five medium caste communities with 8%, 6%, 4% 2% and 1% of firms each, and the remaining firms are assigned to small communities with only two firms each. Table A7 summarizes the externally calibrated parameters, their definitions, and the values assigned to them.

There remain five parameters to be estimated $\Gamma = \{\mu_{lnF}, \sigma_F, \sigma_z, \delta_z, \delta_F\}$. We choose five moments in the data to estimate Γ: the mean and variance of the log number of customers, the variance of network sales, and our estimated effect of castes on the extensive and intensive margins of trade that can be attributed to trading frictions (Table 4), as explained in section 3.2.

In SMM, all moments jointly pin down all unknown parameters. However, there is an intuitive mapping from the targeted moments to model parameters. The mean log number of customers helps identify the mean of the matching costs F whilst the variance of the log number of customers and the variance of network sales identify the variances of productivity and matching costs.

The mapping from our estimates of the effect of castes on trade inferred from Tables 2, 3 and 4 to the model parameters on the role of castes warrants further discussion. Note first that our model is designed to consider how communities facilitate trade by making matches more productive, easier to form or both. Our empirical evidence suggests that a large part of the observed effect of trade can be explained by castes indeed facilitating trade via alleviating frictions in firm-to-firm markets, but not all of it: we also find evidence suggesting that taste-based discrimination against firms from other castes may play a role. Columns 3 and 6 of Table 4 suggest that roughly one-third of the effect of castes on trade remains in contexts with minimal frictions (uncongested courts and homogeneous products). We cannot, of course, rule out that other frictions are still important in those contexts, but this suggests that at least two-thirds of the estimates can be attributed to castes alleviating frictions. We therefore use two-thirds of our preferred estimates of the effect of castes on the intensive and extensive margins of trade (those in the last columns of Tables 2 and 3) as moments in the estimation: 0.66 for the extensive margin and 0.122 for the intensive margin.

Taking logs of our model expression for firm-to-firm sales (11) we see that our specification (1) for the effect of castes on log trade volumes identifies the parameter δ_z if there are no pair-specific unobserved determinants of trade profitability correlated with the likelihood that both trading parties belong to the same caste. Firms which belong to different castes and nevertheless trade may, all else equal, have an above-average realization of such unobserved determinants, so our approach may under-estimate δ_z. Future work will obtain unbiased estimates by implementing the Tobit solution suggested by Crozet et al. (2012). Our estimate of the effect of castes on the extensive margin of trade is similarly the empirical equivalent of our model expression for the share of supplier-client pairs that trade (expression (12)). As this expression makes clear, once δ_z is identified the effect of castes on the extensive margin enables us to also identify the caste-specific matching cost, δ_F.

Overall, our model enables us to estimate the pair-productivity and matching caste parameters that are consistent with the production network moments as well as the reduced-form evidence presented in section 3. Collecting the targeted empirical moments in vector x and the simulated moments in vector $x^s(\Gamma)$, the SMM estimates for Γ solve:

$$\arg\min_{\Gamma}(x - x^s(\Gamma))'(x - x^s(\Gamma)), \tag{13}$$

5.2 Model estimation results

Table 6 reports the results from the SMM estimation. Column 1 reports the value of the targeted moments in the data. Column 2 reports our estimated parameters and value of the simulated moments. We find that the variance of the matching ability is much higher than the variance of productivity, in line with recent evidence on the importance of the firm network for firm outcomes. Crucially, our estimates indicate that communities play a role in both the profitability of the match and the likelihood of the match. We find that $\delta_z = 1.0308$, meaning that within-caste trade is 3% more profitable than trade across-castes. We find that $\delta_F = -2.0953$, implying that matching costs for same-caste pairs are 87% lower (1-exp(-2.09)=0.87) than the across-caste matching costs. Therefore, our estimation suggests a dual role for communities in firm-to-firm networks. Our estimated parameters provide a very good model fit, as all targeted moments are well matched. In addition, the model performs well in untargeted moments. The model predicts a larger average (log) number of same-caste clients (-3.66 relative to -2.50 in the data). We also match the positive correlation between sales per client and firm size (0.04 relative to 0.01 in the data) and a strong negative degree assortativity (slope of -0.28 relative to -0.47 in the data).

Columns 3 and 4 estimate alternative models in which we restrict the effect of communities to operate only through one margin ("No matching effect" eliminates the differences in matching costs, while "No productivity effect" eliminates differences in pair-productivity). As we can see from the targeted and untargeted moments, neither of these restricted models are able to explain the caste-specific moments that we see in the data. The model with the same matching cost within and across-caste overestimates the effect of castes on the intensive margin and does not generate a large enough effect in the same-caste trade probability or in the number of same-caste clients. The model with the same pair-productivity effect within and across-caste cannot match the same-caste effect on sales. These results confirm our finding that communities affect both the intensive margin (trading profitability) and the extensive margin (matching cost).

5.3 Counterfactuals

In this section, we use our estimated model to investigate how communities shape aggregate outcomes such as firm sales, firm connections and welfare. In our empirical section, we found evidence of the role of communities in alleviating contractual frictions in the context of differentiated goods and congested courts. When contractual frictions are high, firm-pairs in the same community are more likely to start a trading relationship and to trade larger values than firms in different castes, highligthing the role of communities as trade facilitators. We perform two counterfactuals to explore the aggregate implications of this trade-facilitating role. In the first counterfactual, we ask: *What would be the aggregate implications of extending the positive effects of castes on trade to* all *potential supplier-client pairs?* To find out, we perform counterfactual 1: all supplier-client pairs enjoy the benefits of within-caste trade: higher trading profitability and lower matching costs. In the second counterfactual, we ask the opposite question: *What would be the aggregate implications of removing the positive effects of castes on trade for all pairs?* To investigate this, we run counterfactual 2: no potential supplier-client pairs benefits from the within-caste higher profitability and lower matching costs.

Panel A of Table 7 reports the results for both counterfactuals in columns 1 and 2. The outcomes are reported relative to the levels of the baseline model, estimated in the previous section. What is the aggregate effect of extending the positive effects of castes on trade to *all* potential supplier-client pairs? We find very large aggregate effects: Welfare in the economy goes up by 32.2%. This comes from a 63.2% increase in total firm-to-firm connections, and an input price reduction of almost 13%. As a result of the increased trading profitability and the lower matching costs, average firm-to-firm trade (network sales) grows by almost 20%. The reason for this large effect is that most potential supplier-client pairs (96.5%) are not same-caste pairs. Thus, extending the positive effects of castes on trade to all potential pairs, we are increasing the trading profitability and the matching ability for almost 97% of firms-pairs in the sample.

Column 2 of panel A reports the results of counterfactual 2: removing the positive effects of castes on trade from all potential supplier-client pairs. This would have more moderate effects, due to the small number of same-caste pairs in the data (3.5%). If we eliminate the higher profitability and lower matching costs that same-caste pairs enjoy, this would reduce welfare by 1%. This reduction would come from 2.3% lower connections with clients and 2% lower average firm-to-firm sales.

Finally, in panel B of Table 7 , we provide a decomposition of the effects of counterfactual 1 (extending the positive effects of castes on trade) of each of the margins separately. We find that most of the effect comes from the reduction in the matching cost for all different-caste firm pairs: keeping differences in profitability but reducing the matching cost for all pairs would provide 100% of the increase in sales and network connections, and 75% of the welfare gains. On the contrary, providing all pairs with the extra 3% in trade profitability but keeping the different matching cost would have almost no effects on aggregate variables.

5.4 Distributional effects

Finally, we investigate the distributional implications of our two counterfactuals for firm connections and firm-to-firm sales. Our first object of interest is the firm network. Figure 3 plots the distribution of the trade probability (number of clients normalised by number of firms) in the three different scenarios: baseline, extension of caste positive trade effects (C1) and removal of caste positive trade effects (C2). Relative to the baseline, extending the positive effect of castes on trade shifts the distribution to the right, with a significant drop in the number of firms with very few connections, and a higher density in all other bins. These patterns shows how reducing contractual frictions would allow small firms to gain more clients. On the other hand, removing the positive effect of castes on trade would not alter substantially the client distribution.

The second outcome of interest in the firm size distribution. There is a wide literature highlighting the lack of middle-sized and large firms in the data in LMICs. Our second counterfactual illustrates the crucial role that firm-to-firm networks could play in increasing firm size. As we can see in figure 4, the firm size distribution shifts to the right when we allow all firms to enjoy the positive effect of castes on trade (C1). On the contrary, removing the positive effect of castes on trade for all pairs (C2), would not affect the firm-size distribution substantially, due to the few number of same caste supplier-client pairs.

Finally, we look at the distributional impact of counterfactual 1 (extension of caste positive trade effects to all firm-pairs), relative to firm size at baseline. Figure 5 plots the change in average firm-to-firm trade for each firm after extensding the postive effect of castes on trade effects to all potential firm-pairs, plotted against the firm size (total sales) at baseline. The figure shows that sales increase for almost all firms, except for the very large ones. In particular, the smallest firms at baseline gain the most sales, achieving increases of up to 60% in sales to other firms. These counterfactuals point to potentially large effects of improving firm-to-firm networks for development outcomes such as firm size and welfare. Using our model estimates, we find large aggregate effects of same-caste trading. In order to design policies that can provide the benefits highlighted by our counterfactuals, future work should be directed towards a better understanding of the reasons and mechanisms through which communities facilitate trade.

6 Conclusion

This paper considers the role of community (caste) networks in shaping firm-to-firm trade in India. Using panel data on firm-to-firm transactions and information on the firm owner's community, we find that two firms are twice as likely to trade, and when they do trade, trade 20% more, when they belong to the same community. We provide evidence consistent both with communities alleviating frictions and with taste-based discrimination.

To understand the aggregate effects of communities on the economy via trade, we build a model of network formation in which communities affect both the productivity of a client-supplier relationship and the cost of forming the relationship. Estimating model parameters using our reduced-form evidence, we find that extending the positive effects of communities on trade to all potential supplier-client relationships would increase the number of trading relationships by 60%, and lead to firm growth, particularly amongst smaller firms.

References

Alesina, A. and E. L. Ferrara (2005): "Ethnic diversity and economic performance," *Journal of economic literature*, 43, 762–800.

Alfaro-Ureña, A., I. Manelici, and J. P. Vasquez (2022): "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages," 137, 1495–1552.

Anderson, S. (2011): "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics*, 3, 239–63.

Antràs, P., T. C. Fort, and F. Tintelnot (2017): "The Margins of Global Sourcing: Theory and Evidence from US Firms," 107, 2514–2564.

Ash, E., S. Asher, A. Bhowmick, S. Bhupatiraju, D. Chen, T. Devi, C. Goessmann, P. Novosad, and B. Siddiqi (2021): "In-group bias in the Indian judiciary: Evidence from 5 million criminal cases," Working Paper.

Atalay, E., A. Hortaçsu, M. J. Li, and C. Syverson (2019): "How wide is the firm border?" *The Quarterly Journal of Economics*, 134, 1845–1882.

Atkin, D. and A. K. Khandelwal (2020): "How Distortions Alter the Impacts of International Trade in Developing Countries," *Annual Review of Economics*, 12, 213–238.

Atkin, D., A. K. Khandelwal, and A. Osman (2017): "Exporting and Firm Performance: Evidence from a Randomized Experiment*," 132, 551–615.

Bandyopadhyay, S., C. C. Coughlin, and H. J. Wall (2008): "Ethnic Networks and US Exports*," *Review of International Economics*, 16, 199–213.

Banerjee, A., E. Duflo, M. Ghatak, and J. Lafortune (2013): "Marry for What? Caste and Mate Selection in Modern India," *American Economic Journal: Microeconomics*, 5, 33–72.

Banerjee, A. and K. Munshi (2004): "How Efficiently Is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur," *The Review of Economic Studies*, 71, 19–42.

Banerjee, A. V. and E. Duflo (2000): "Reputation effects and the limits of contracting: A study of the Indian software industry," *The Quarterly Journal of Economics*, 115, 989–1017.

Becker, G. S. (1957): *The economics of discrimination*, University of Chicago press.

Bernard, A. B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes (2022):

"The Origins of Firm Heterogeneity: A Production Network Approach," 130, 1765–1804.

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

Bernard, A. B., A. Moxnes, and K. H. Ulltveit-Moe (2018): "Two-Sided Heterogeneity and Trade," *The Review of Economics and Statistics*, 100, 424–439.

Boehm, J. and E. Oberfield (2020): "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *The Quarterly Journal of Economics*, 135, 2007–2058.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019a): "Inaccurate Statistical Discrimination: An Identification Problem," NBER Working Papers 25935, National Bureau of Economic Research, Inc.

Bohren, J. A., A. Imas, and M. Rosenberg (2019b): "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 109, 3395–3436.

Caria, S. and J. Labonne (2022): "Village Social structure and Labor Market PerformanceEvidence from the Philippines," Tech. rep., Mimeo, University of Warwick.

Cassan, G. (2020): "The Economics of Caste," Tech. rep., Chapter prepared for the Oxford Handbook of Caste in Contemporary/Modern Times, Surinder S. Jodhka and Jules Naudet (editors).

Cassan, G., D. Keniston, and T. K. Kleineberg (2021): "A Division of Laborers : Identity and Efficiency in India," Policy Research Working Paper Series 9544, The World Bank.

Ciani, A., M. C. Hyland, N. Karalashvili, J. L. Keller, A. Ragoussis, and T. T. Tran (2020): *Making It Big: Why Developing Countries Need More Large Firms*, The World Bank.

Combes, P.-P., M. Lafourcade, and T. Mayer (2005): "The trade-creating effects of business and social networks: evidence from France," 66, 1–29.

Crozet, M., K. Head, and T. Mayer (2012): "Quality Sorting and Trade: Firm-level Evidence for French Wine," *Review of Economic Studies*, 79, 609–644.

Dai, R., D. Mookherjee, K. Munshi, and X. Zhang (2020): "The Community Origins of Private Enterprise in China," Boston University - Department of Economics - The Institute for Economic Development Working Papers Series dp-320, Boston University - Department of Economics.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, Markups, and Trade Reform," 84, 445–510.

Fafchamps, M. (2000): "Ethnicity and credit in African manufacturing," *Journal of Development Economics*, 61, 205–235.

Fisman, R., D. Paravisini, and V. Vig (2017): "Cultural Proximity and Loan Outcomes," *American Economic Review*, 107, 457–92.

Fisman, R. J. (2003): "Ethnic Ties and the Provision of Credit: Relationship-Level Evidence from African Firms," *The B.E. Journal of Economic Analysis & Policy*, 3, 1–21.

Gadenne, L., T. Nandi, and R. Rathelot (2022): "Taxation and Supplier Networks: Evidence from India," Tech. rep., CEPR Discussion Paper 13971.

Gould, D. M. (1994): "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," 76, 302.

Greif, A. (1993): "Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition," *American Economic Review*, 83, 525–48.

——— (2006): *Institutions and the path to the modern economy: Lessons from medieval trade*, Cambridge University Press.

Guiso, L., P. Sapienza, and L. Zingales (2009): "Cultural biases in economic exchange?" *The Quarterly Journal of Economics*, 124, 1095–1131.

Hansman, C., J. Hjort, G. León-Ciliotta, and M. Teachout (2020): "Vertical Integration, Supplier Behavior, and Quality Upgrading among Exporters," *Journal of Political Economy*, 128, 3570–3625.

Head, K. and T. Mayer (2014): "Gravity Equations: Workhorse,Toolkit, and Cookbook," in *Handbook of International Economics*, Elsevier, vol. 4, 131–195.

Hjort, J. (2014): "Ethnic divisions and production in firms," *The Quarterly Journal of Economics*, 129, 1899–1946.

——— (2021): "Ethnic Investing and the Value of Firms," CEPR Discussion Papers 16316, C.E.P.R. Discussion Papers.

Hsieh, C.-T. and B. A. Olken (2014): "The Missing Missing Middle," *Journal of Economic Perspectives*, 28, 89–108.

Huneeus, F. (2018): "Production network dynamics and the propagation of shocks," .

Iyer, R. and A. Schoar (2015): "Ex post (in) efficient negotiation and breakdown of trade," *American Economic Review*, 105, 291–94.

La Ferrara, E. (2002): "Self-help Groups and Income Generation in the Informal Settlements of Nairobi," *Journal of African Economies*, 11, 61–89.

Liao, S., I. S. Kim, S. Miyano, and H. Zhang (2020): *concordance: Product Concordance*, r package version 2.0.0.

Macchiavello, R. (2022): "Relational Contracts and Development," *Annual Review of Economics*, 14, 337–362.

Macchiavello, R. and A. Morjaria (2015): "The value of relationships: evidence from a supply shock to Kenyan rose exports," *American Economic Review*, 105, 2911–45.

Mazzocco, M. and S. Saini (2012): "Testing efficient risk sharing with heterogeneous risk preferences," *American Economic Review*, 102, 428–68.

McMillan, J. and C. Woodruff (1999): "Interfirm relationships and informal credit in Vietnam," *The Quarterly Journal of Economics*, 114, 1285–1320.

Munshi, K. (2011): "Strength in numbers: Networks as a solution to occupational traps," *The Review of Economic Studies*, 78, 1069–1101.

——— (2014): "Community Networks and the Process of Development," *Journal of Economic Perspectives*, 28, 49–76.

——— (2019): "Caste and the Indian Economy," *Journal of Economic Literature*, 57, 781–834.

Munshi, K. and M. Rosenzweig (2006): "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy," *American Economic Review*, 96, 1225–1252.

——— (2016): "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 106, 46–98.

Nunn, N. (2007): "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade," *Quarterly Journal of Economics*, 122, 569–600, reprinted in D. Bernhofen (ed.), Empirical International Trade, Edward Elgar Publishing, 2010.

Rauch, J. E. (1999): "Networks versus markets in international trade," *Journal of international Economics*, 48, 7–35.

——— (2001): "Business and Social Networks in International Trade," 39, 1177–1203.

Rauch, J. E. and V. Trindade (2003): "Information, International Substitutability, and Globalization," 93, 775–791.

Singh, K. (1996): *Communities, Segments, Synonyms, Surnames and Titles*, National series, Anthropological Survey of India.

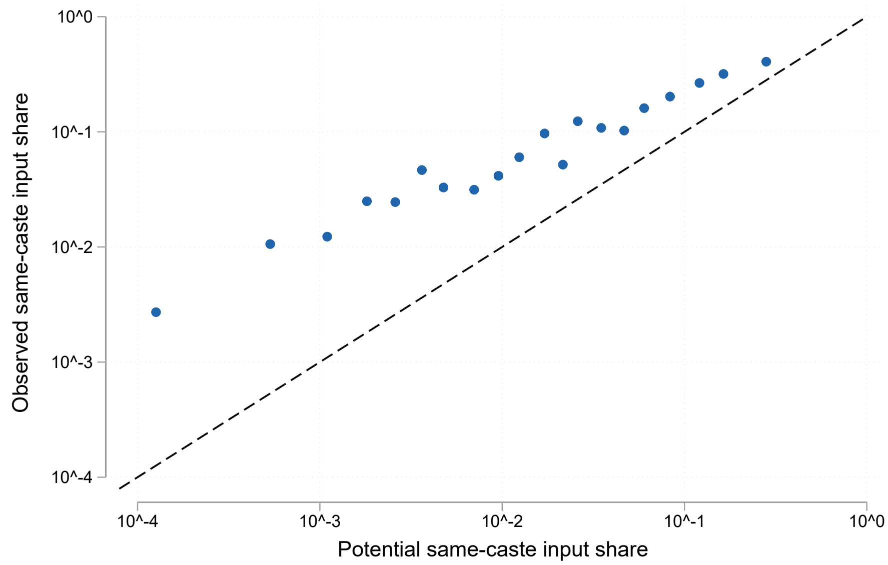
Startz, M. (2021): "The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade," Tech. rep.

Weber, A. and C. Zulehner (2014): "Competition and gender prejudice: Are discriminatory employers doomed to fail?" *Journal of the European Economic Association*, 12, 492–521.

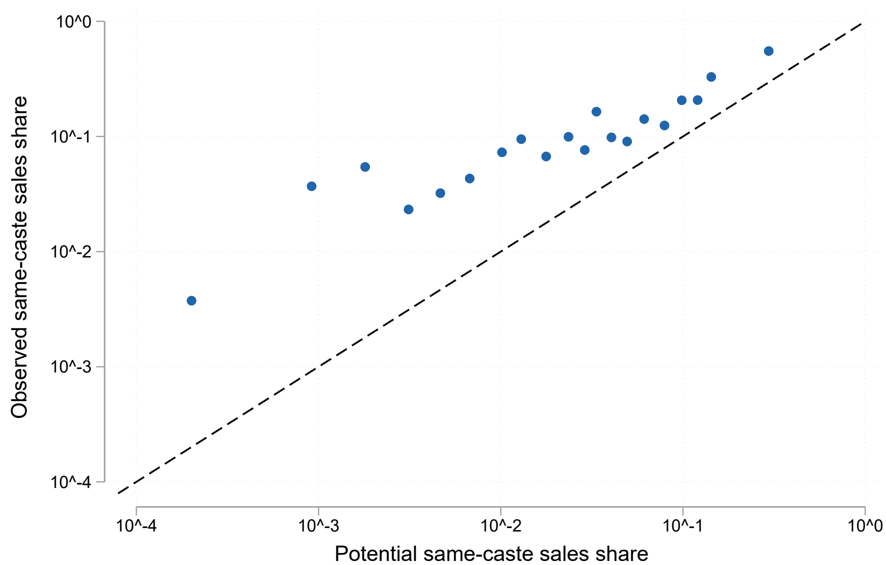
Woodruff, C. (2002): "Non-contractible investments and vertical integration in the Mexican footwear industry," *International Journal of Industrial Organization*, 20, 1197–1224.

Figure 1: Potential and observed same-caste trade

(a) Inputs



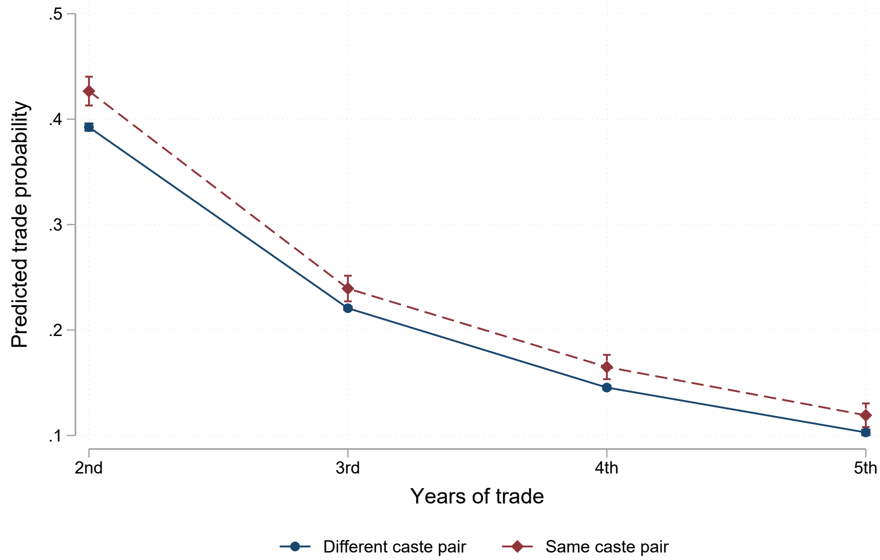
(b) Sales



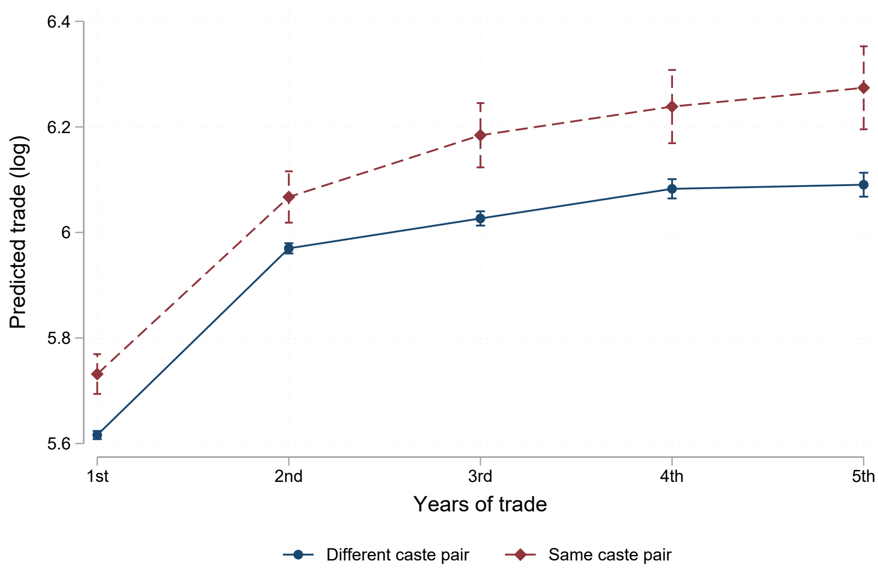
The graphs illustrate the relationship between potential same-caste trade and observed same-caste trade, averaged within 20 vintiles of potential own-caste trade share. Potential same-caste trade (input or sales) share is the average share of same-caste trading partners in a firms' all potential trading partners, where each partner is weighted by its average network sales. Panel a) plots firms' observed input share purchased from same-caste suppliers as a function of their potential same-caste input share. Panel b) plots firms' observed sales share sold to same-caste clients as a function of their potential same-caste sales share. Each firm is weighted by its average annual network trades, we exclude the 5 largest firms.

Figure 2: Predicted trade, new pairs 2011-2012

(a) Predicted trade probability

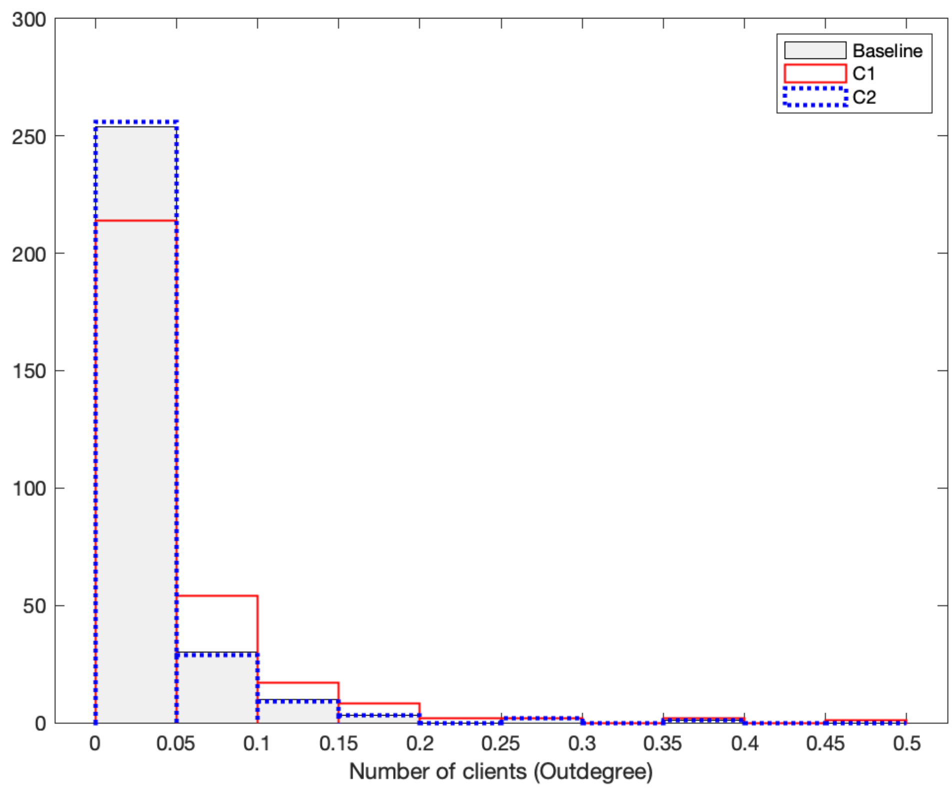


(b) Predicted trade volume (conditional)



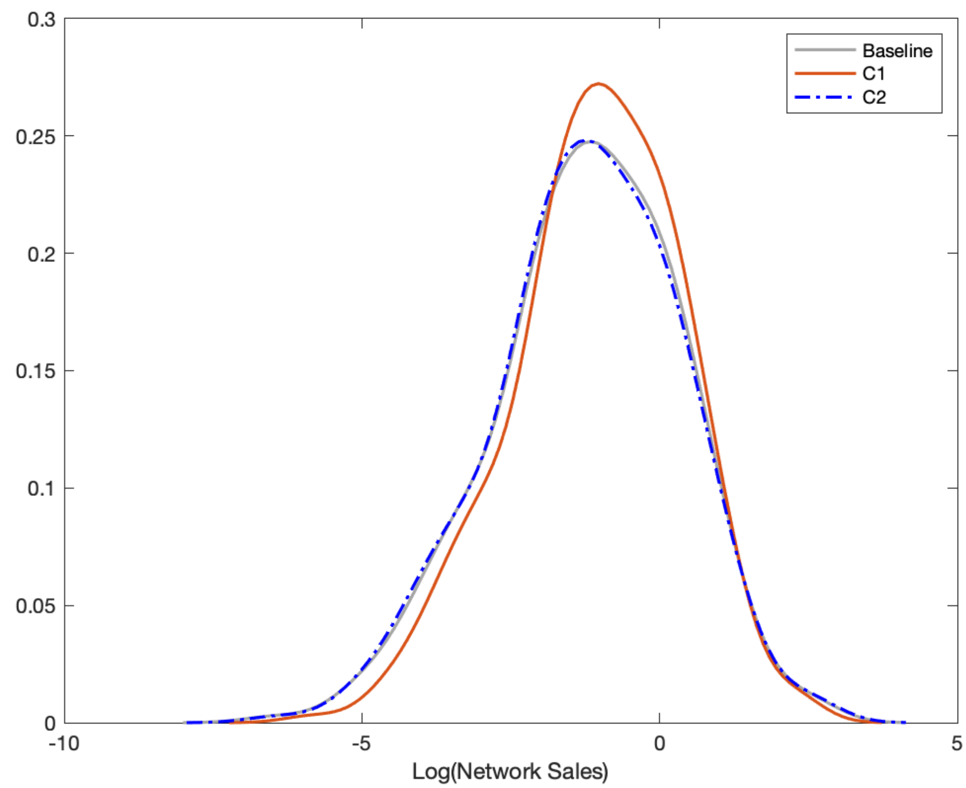
These graphs plot the trade dynamics for newly formed trade relationships separately for same-caste relationships and all other relationships. Each point represents a predicted outcome, computed separately for pairs for which both firm owners belong to the same caste and for other pairs, with 95% confidence intervals. Our sample is the sample of firm pairs that trade in the second year in our data (2011-12) but not in the previous year, and for which both the supplier and the supplier are present in the data in all subsequent years. Estimates of the effect of the two firms being in the same caste on trade over time are obtained by using an augmented version of our specification 1 that allows the same caste effect to vary in each year: $Y_{ijt} = \sum_k \beta^k \mathbb{1}(c_i = c_j) \times \mathbb{1}(Year\_trade = k) + \sum_{k=1}^{k=5} \theta^k \mathbb{1}(Year\_trade = k) + \gamma X_{ijt} + \mu_i + \mu_j + \epsilon_{ijt}$. In panel (a), Y_{ijt} is an indicator equal to 1 if firm pair ij trades in year t; in panel (b), Y_{ijt} is the log of observed trade between i and j in year t. We control throughout for supplier and client fixed effects, firms' joint locations and products sold, following the specification used in column (4) of Tables 2 and 3. Indicated confidence intervals at the 95% level, standard errors are clustered two-way at the supplier and client levels. Sample size: (a) N=485,324 (b) N=226,709.

Figure 3: Distribution of the number of network links



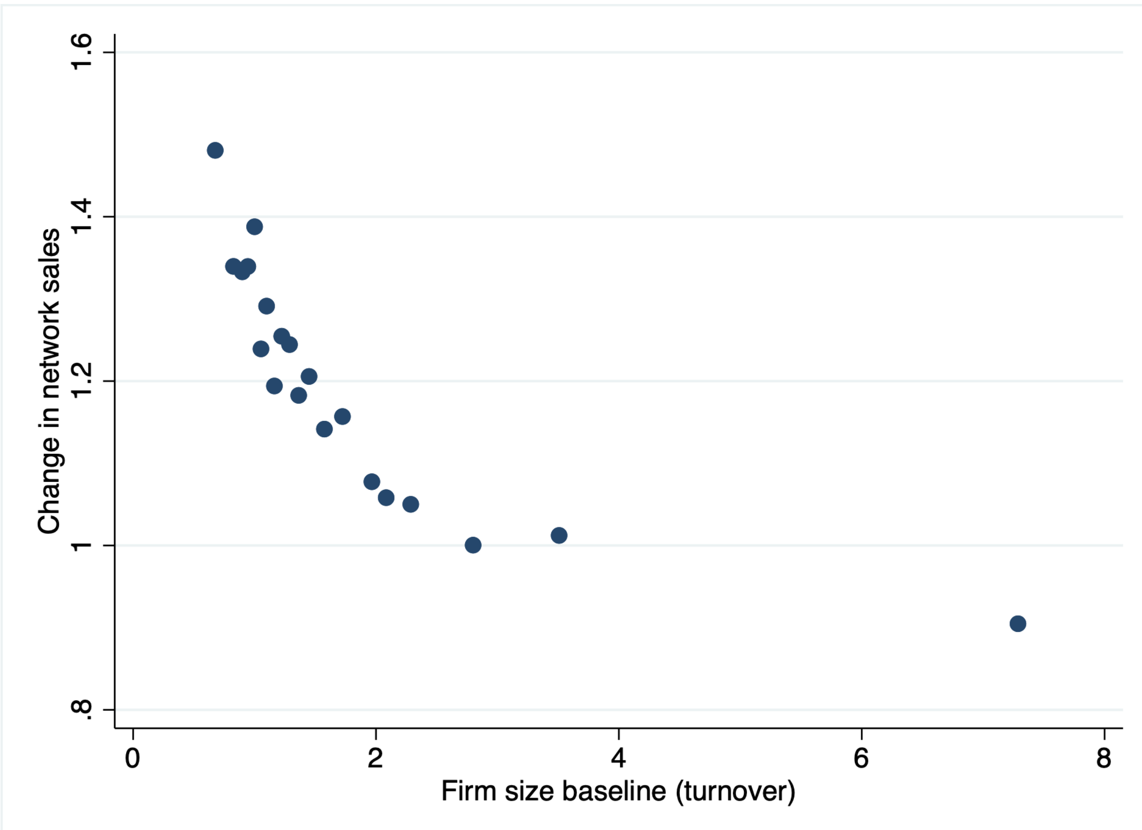
The figure plots the distribution of the total number of clients per firm in the estimated model ('Baseline') and in two counterfactual scenarios. Counterfactual 1 ('C1') assumes all firm-pairs can benefit from the positive effect of castes on trade . Counterfactual 2 ('C2') assumes that we remove the positive effect of castes on trade for all pairs, so no firms can benefit from the higher pair-productivity or matching ability within-caste.

Figure 4: Firm size distribution



The figure plots the distribution of (log) firm size in the estimated model (baseline) and in two counterfactual scenarios. Counterfactual 1 ('C1') assumes all firm-pairs can benefit from the positive effect of castes on trade . Counterfactual 2 ('C2') assumes that we remove the positive effect of castes on trade fow all pairs, so no firms can benefit from the higher pair-productivity or matching ability within-caste.

Figure 5: Change in sales and firm size



The binned scatterplots group firms into 20 equal-sized bins by firm size (log), and compute the mean of the variables on the x- and y-axes in each bin. Change in network sales is network sales under counterfactual 1 divided by network sales in the baseline model. Counterfactual 1 ('C1') assumes all firm-pairs can benefit from the positive effect of castes on trade.

Table 1: Sample descriptives

	Mean	SD	Median
A: Firms			
Turnover (1000 INR)	21.41	245.52	3.30
Years active	3.95	1.91	4
Is supplier	0.62	0.49	1
Is client	0.92	0.28	1
# Clients	2.98	9.89	1
# Suppliers	3.01	4.06	2
# Potential clients	508.22	795.15	257.67
# Potential suppliers	443.02	461.09	285.67
# Potential same-caste clients	17.74	41.43	1.67
# Potential same-caste suppliers	15.51	30.63	4.33
Observations		106,775	

	Mean	SD	Median
B: Potential trade			
Trade probability (%)	0.72	8.46	0
Transaction amount (1000 INR)	14.93	10780.95	279.38
Same caste probability (%)	3.57	18.55	0
Observations		202,495,680	

Panel A presents firm-level annual summary statistics.The number of potential suppliers and clients are constructed using the potential trade sample described in the text. 'Years active' is the number of fiscal years in which the firm is observed in our data. 'Is supplier' ('Is client') is an indicator equal to one if a firm is observed at least once selling to (buying from) another firm in our sample. In Panel B each observation is a potential supplier-client pair in a given year, the sample is the potential trade sample described in the text, and the transaction amount is conditional on the pair trading.

Table 2: Caste effect on the extensive margin of trade

	(1) $\mathbb{1}(Trade)$	(2) $\mathbb{1}(Trade)$	(3) $\mathbb{1}(Trade)$	(4) $\mathbb{1}(Trade)$
Same caste	1.312 ***	1.003***	1.000***	0.990***
SE	(0.029)	(0.026)	(0.021)	(0.021)
Client location X Supplier location FE		X	X	X
Client X Year FE			X	X
Supplier X Year FE			X	X
Client Product X Supplier Product FE				X
Obs. (thousand)	202,496	202,496	202,496	202,496

This table presents results obtained from running specification (1) on our potential trade sample, described in the text. The variable 'same caste' is an indicator equal to 1 divided by the mean probability of trade in the sample (.00726) if the two firms are in the same caste, 0 otherwise, so that coefficients can be read as the effect of caste on the probability that two firms trade. All columns from column 2 onwards include fixed effects for each interaction of the location (postcode) of the supplier and the location (postcode) of the client, columns (3) and (4) include supplier $\times$ year and client $\times$ year fixed effects and column (4) includes fixed effects for each interaction of the product sold by the supplier and the product sold by the client. Standard errors clustered two-way at the level of the supplier and the client in parentheses.Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 3: Caste effect on the intensive margin of trade

	(1) Log. trade	(2) Log. trade	(3) Log. trade	(4) Log. trade
Same caste	0.336***	0.332***	0.198***	0.183***
	(0.016)	(0.012)	(0.009)	(0.009)
Client location X Supplier location FE		X	X	X
Client X Year FE			X	X
Supplier X Year FE			X	X
Client Product X X Supplier Product FE				X
Obs. (thousand)	1,461	1,461	1,461	1,461

This table presents results obtained from running specification (1), described in the text on the sample of all trading supplier-client pairs. The outcome variable is log trade within the pair in the year. The variable 'same caste' is an indicator equal to 1 if the two firms are in the same caste, 0 otherwise. All columns from column 2 onward include fixed effects for each interaction of the location (postcode) of the supplier and the location (postcode) of the client, columns (3) and (4) include supplier $\times$ year and client $\times$ year fixed effects and column (4) includes fixed effects for each interaction of the product sold by the supplier and the product sold by the client. Standard errors clustered two-way at the level of the supplier and the client in parentheses.Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 4: Caste and trade frictions

	(1) $\mathbb{1}(Trade)$	(2) $\mathbb{1}(Trade)$	(3) $\mathbb{1}(Trade)$	(4) Log. trade	(5) Log. trade	(6) Log. trade
Same caste	0.619***	0.739***	0.363***	0.123***	0.116***	0.056*
	(0.048)	(0.067)	(0.080)	(0.025)	(0.021)	(0.031)
Same caste X Court cong.	0.626***		0.629***	0.099**		0.098**
	(0.074)		(0.074)	(0.039)		(0.039)
Same caste X Rel. spec.		0.394***	0.396***		0.104***	0.104***
		(0.095)	(0.095)		(0.029)	(0.029)
Obs. (thousand)	202,496	202,496	202,496	1,461	1,461	1,461

This table presents results obtained from running specification (1) augmented with interaction terms. In columns (1) to (3) the dependent variable is an indicator equal to 1 if the two firms trade in the year and the sample is the potential trade sample defined above. In columns (4) to (6) the dependent variable is the log of trade between the two firms in the year and the sample is restricted to all trading pairs. The variable 'same caste' is an indicator equal to 1 if the two firms are in the same caste, 0 otherwise. The variable court congestion is the share of filed cases which were decided cases after 2 years in the client's district. The variable 'relationship-specificity' measures the share of goods in the supplier's NIC4 category that are not traded on central exchanges or with a reference price according to Rauch (1999). In columns (1), (2), (3), the variables 'same caste' and all of the interaction terms are divided by the mean probability of trade in the sample (.00726). All columns include fixed effects for each interaction of the location (postcode) of the supplier and the location (postcode) of the client, supplier × year and client × year fixed effects and fixed effects for each interaction of the product sold by the supplier and the product sold by the client. Standard errors clustered two-way at the level of the supplier and the client in parentheses.Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table 5: Caste effect on firm exit

	(1) Exit	(2) Exit	(3) Exit	(4) Exit
Own caste share, inputs (demeaned)	0.095*** (0.023)	0.091*** (0.023)	0.083** (0.025)	0.085** (0.031)
Age	-0.081*** (0.002)	-0.087*** (0.002)	-0.086*** (0.002)	-0.086*** (0.003)
Log turnover	-0.275*** (0.005)	-0.282*** (0.005)	-0.315*** (0.007)	
Stratification		Postcode	Product & Postcode	Product & Postcode & Size decile
Observations	364,967	364,967	364,967	364,967

This table presents the coefficients from our Cox model estimating the probability of firm exit, as described in the text. The coefficient 'own-caste share' is the share of input purchased from firms from the client's caste in a given year, demeaned by the industry average. Firm age is the time in years since registration. Stratification lists the variables that define the subgroups for comparison, i.e. subgroup-specific baseline. Standard errors are clustered at the Postcode X good level. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 6: Estimation results: parameter estimates and model fit

	Data (1)	Baseline (2)	No matching effect (3)	No productivity effect (4)
Estimated Parameters				
μ_{lnF}		15.2629	15.1432	15.2645
σ_z		0.0652	0.0382	0.0661
σ_F		3.4693	3.4781	3.4688
δ_z (caste productivity)		1.0308	1.1189	
δ_f (caste matching cost)		-2.0953		-2.1943
Targeted moments				
var(sales)	**2.380**	2.380	2.337	2.3812
var(ln n clients)	**2.200**	2.200	2.200	2.1989
mean(ln n clients)	**-4.500**	-4.500	-4.484	-4.4998
RF coeff on sales	**0.110**	0.110	0.336	0.0171
RF coeff on trade prob	**0.600**	0.600	-0.134	0.6026
Untargeted moments				
ln n caste clients	**-2.500**	-3.668	-4.358	-3.664
Market sh & n clients	**0.011**	0.047	0.042	0.047
Neg degree assortativity	**-0.467**	-0.281	-0.267	-0.282

This table presents the results from the SMM estimation. The firms panel reports the values of the estimated parameters. The second and third panels report the values of targeted and untargeted moments. Sales are demeaned, Log number of clients ("ln n clients") is normalised by the total number of firms. "RF coeff on sales" stands for the coefficient on the *same caste* dummy variable estimated in equation 1 using log sales as the dependent variable, while "RF coeff on trade prob" stands for the coefficient on the *same caste* dummy variable estimated in equation 1 using a *Trade* dummy to measure the probability of trading. Both coefficients are scaled down to 0.6 to account only for the effect of castes in the presence of high contractual frictions. "ln n caste clients" is the number of same caste clients normalised by the potential number of same caste clients for each firm. "Market sh & n clients" is the value of the slope of regressing the average market share that a firm has on the number of clients. "Neg degree assortativity" is the slope of regressing the average connectivity of a firm's clients (number of suppliers of each clients), over the number of clients (as reported in Figure A6). Column 2 reports the results from the baseline model while column 3 reports the results from a restricted model with only an intensive margin caste parameter (the productivity-pair effect), while column 4 reports the results from a restricted model with only the extensive margin parameter (matching cost effect).

Table 7: Aggregate effects of Community networks: counterfactual exercises

A: Counterfactuals	C1	C2
Welfare change	1.3235	0.99005
Change in total number of connections	1.6322	0.97753
Change in average network sales	1.1947	0.98052
Change in average input price	0.87241	1.0045

B: Counterfactual 1 decomposition	Matching cost effect	Pair-productivity effect
Welfare change	1.2646	1.0361
Change in total number of connections	1.6322	0.97753
Change in average network sales	1.1947	0.98052
Change in average input price	0.89309	0.98385

This table presents the results from the counterfactual exercises. Panel A reports the results from our two main counterfactuals. C1 reports the results from a counterfactual exercise in which we extend the positive effect of caste on trade to all potential supplier-client pairs. C2 reports the results from the second counterfactual in which we remove all the positive effects of castes on trade, both the higher productivity-pair effect and the lower matching costs. Panel B presents the results from decomposing counterfactual C1 into the effect of the matching cost parameter and the effect of the pair-productivity parameter. Column 1 in panel B presents the counterfactual outcomes of extending only the extensive margin caste benefit: all supplier-client pairs enjoy the lowest matching costs that we observe within communities, while the pair-productivity effect is removed. Column 2 in panel B reports the outcomes of extending only the intensive margin caste benefit: all supplier-client pairs enjoy the higher pair-productivity that we observe within communities, while the matching cost effect is removed: all supplier-client pairs enjoy the higher profitability of same-caste pairs, but the matching cost is differences are removed.

Appendix

A Data construction

We use the data by Cassan et al. (2021) based on the data collected by Singh (1996), which lists 2,205 castes (or 'main communities') and information on the names of their various associated subgroups, synonyms, surnames, and the respective sources or origin. In our illustrative example in Figure A1 the castes is Gareri. We construct a list of all names and their associated castes. Since a name can be listed for more than one caste, we often find multiple castes for a given name and order them alphabetically. In our assignment, we only keep the most relevant matches by origin and type. In step 1, we only match castes to names which are listed as surnames with origin West Bengal (from the example Figure A1, the name Bhagat would be matched with the caste Gareri). In step 2, we match all remaining unmatched names to castes based on other group names from West Bengal (e.g., we would match the caste Gareri to the name Goneri *if* we could not find another caste in step 1). We repeat this procedure twice more, matching unmatched names using all surnames from other origins (step 3), and using all other group names from other sources (step 4).

We match the names on our list to the names in the VAT records; 82% of firms have a name recorded in the original data set. We perform fuzzy string matching to account for different spellings and manually verify every match, which accounts for 8% of our matches. Overall, we find a caste for 91 % of firms with a name in the VAT data [13], or 75% of all firms. From this sample, we are pruning based on the trade network and only keep firms which are trading with at least one firm with an assigned caste, dropping an additional 19% of firms.

In our final data, we assign 66 % of firms to castes based on surnames with origin West Bengal and 20% based on surnames from other states[14]. 65% of firms have more than one caste assigned to them, 40% have five or more. We perform two robustness checks on caste matching: first using only the matches based on West Bengali surnames and second, if available, using the second assigned caste in alphabetical order (Table A3 and Table A4). Considering the distribution of

[13] including firm names or other non-surnames

[14] We furthermore assign 6% based on other group names from West Bengal, and 8% based on group names from other states

firms in our final sample across size bins, districts, and sectors, we find that it does not seem to differ much in observables from the universe of firms in the VAT data: we lose proportionally more firms at the extreme ends of the size distribution, mostly very small firms, and we see very little systematic deviation in the geographic and sectoral distribution (Table A1).

Figure A1: Example entry for the caste 'Gareri' from Singh (1996)

GARERI

Synonyms: Bherihar, Pal [Bihar]
Goneri, Gonrhi [West Bengal]

Groups/subgroups: Dhengar, Gangojoli, Nikhar, Phurukbadi [Bihar]
Dhangarh, Nikhar [West Bengal]

* *Subcastes:* Dhengar, Farakhabadi, Gangajali, Nikhar [H.H. Risley]

Titles: Kamblia, Kammali, Marar, Ratu [H.H. Risley]

Surnames: Bhagat, Chowdhury, Mandal, Pal [Bihar]
Bhagat, Choudhury, Ghosh, Pal [West Bengal]

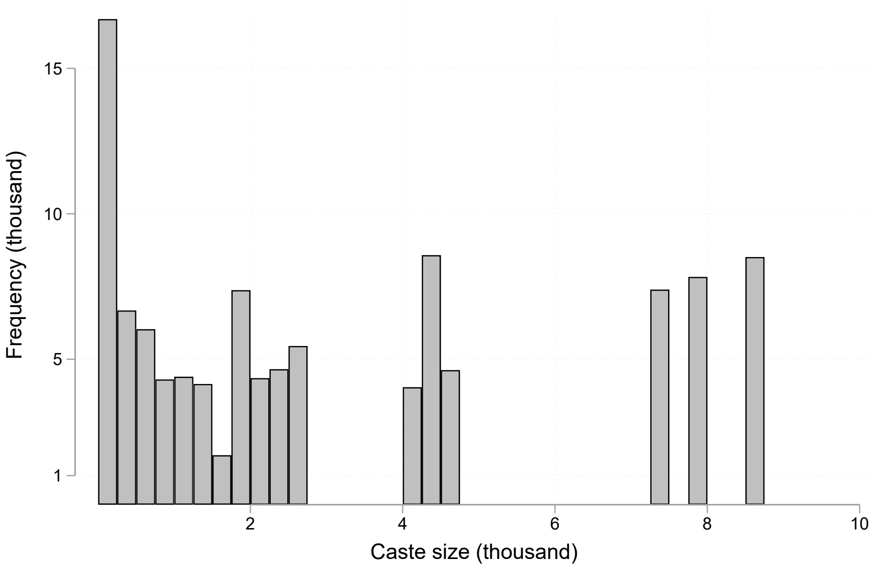
Exogamous units/clans: Ahir, Bandharia, Chowharia, Khandel [West Bengal]

Exogamous units/clans (gotra): Ahir, Basdharia, Bilar, Chandel, Chaurasia, Nakwar [Bihar]

Exogamous units/lineages: Ahir, Bandharia, Chowharia, Khandel [West Bengal]

B Additional Figures and Tables

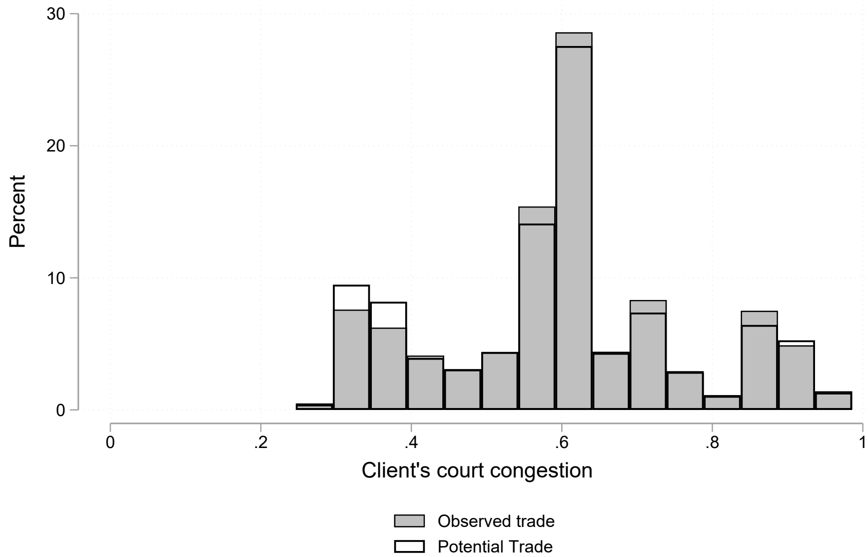
Figure A2: Distribution of caste sizes



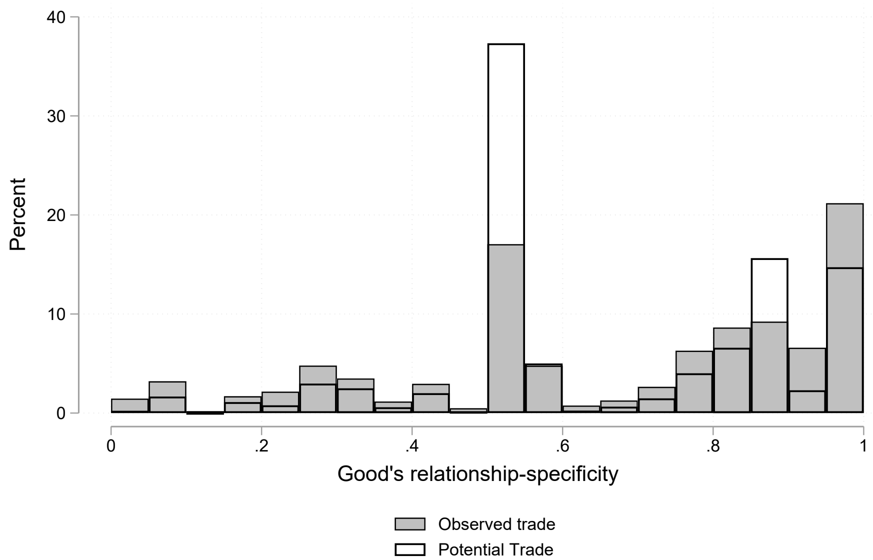
This graph plots the size distribution of every firm's assigned caste. Size refers the total number of firms in our sample which are assigned to a given caste. Bars can represent more than one caste.

Figure A3: Distribution of trade frictions

(a) Client's court congestion, pair-level

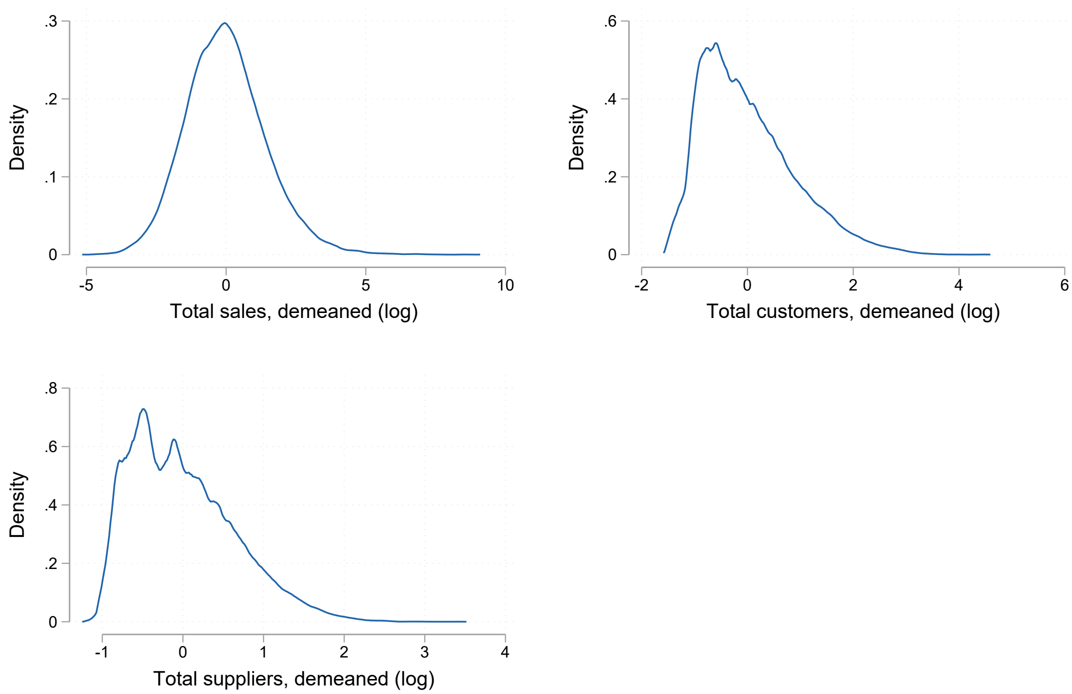


(b) Supplier's product relationship-specificity, pair level



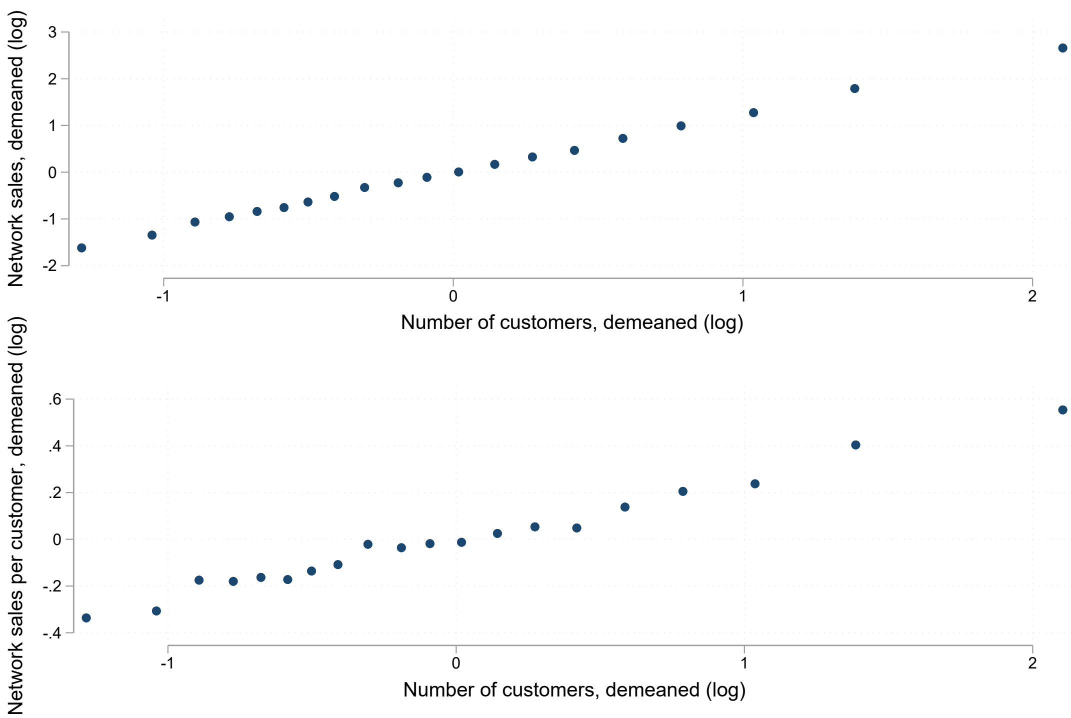
This graph illustrates the distribution of the trade friction variables which are used and described in section X. Panel (a) plots the distribution of the court congestion variable, on the (potential) client's side, for all pairs in our observed (potential) trade data. Panel (b) plots a histogram relationship-specificity of the (potential) supplier's good for all pairs in our observed (potential) trade data.

Figure A4: Distribution of firm sales, number of customers and number of suppliers.



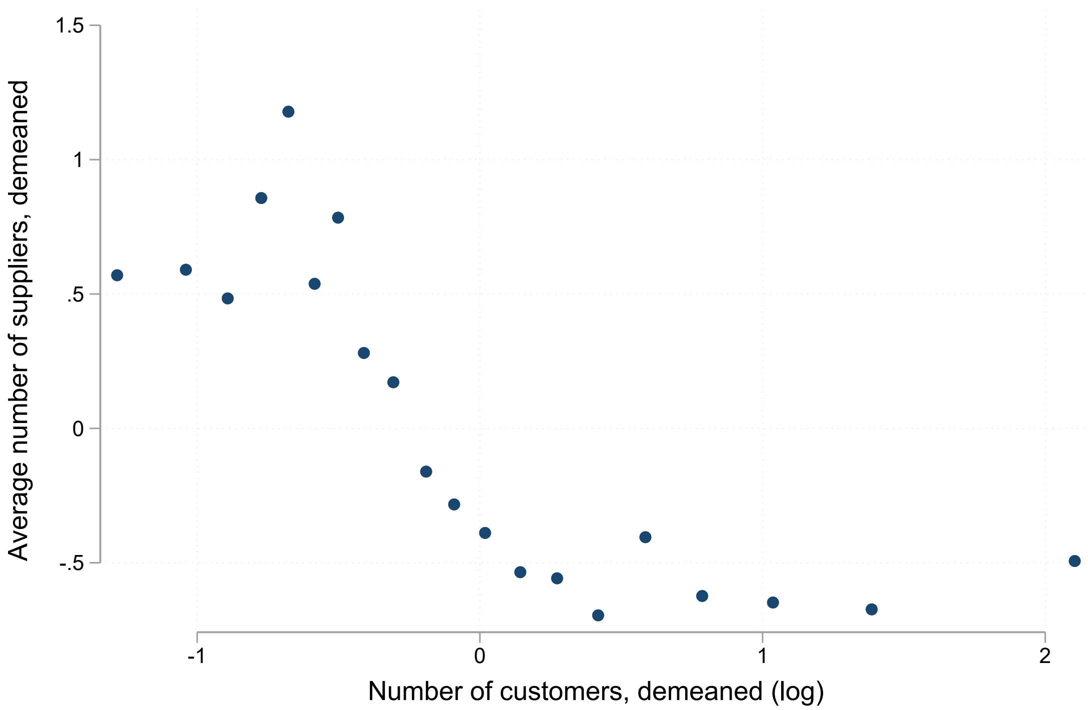
The graph illustrates the density of network sales, total customers and total suppliers in the firms in our dataset.

Figure A5: Total Network Sales, Average Sales and Number of Customers



The binned scatterplots group firms into 20 equal-sized bins by number of customers (log), and compute the mean of the variables on the x- and y-axes in each bin. Network sales are firm's total sales to customers in the domestic production network. All variables are demeaned by 4-digit industry averages. The upper panel plots networks sales over total number of customers while the lower panel plots average sales per customer over total number of customers.

Figure A6: Degree Assortativity



The binned scatterplot groups firms into 20 equal-sized bins by number of customers (log), and computes the mean of the variables on the x and y-axes in each bin. Average number of suppliers refers to the geometric mean of the number of suppliers serving the customers of firm i. All variables are demeaned by 4-digit industry averages.

Table A1: Final sample after cleaning

	VAT data	Final sample
Turnover (INR):		
$\leq$ 500	16.71	10.82
500.1 - 1000	11.40	11.18
1,000.1 - 5,000	35.29	39.66
5000.1 - 10,000	12.05	13.87
10,000.1-100,000	20.21	21.46
$>$ 100,000	4.34	3.02
District:		
Bankura	1.44	1.34
Bardhaman	6.31	6.41
Birbhum	2.14	1.98
Coochbehar	1.40	1.41
Dakshin Dinajpur	0.55	0.53
Darjeeling	4.04	4.67
Hoogly	4.52	4.72
Howrah	11.54	13.89
Jalpaiguri	2.42	2.46
Kolkata	42.63	39.64
Malda	1.36	1.36
Medinipur (E)	2.59	2.37
Medinipur (W)	2.43	2.37
Murshidabad	2.24	1.96
Nadia	2.63	2.62
North 24 Parganas	5.69	6.00
Others	0.02	0.02
Purulia	1.14	1.07
South 24 Parganas	3.96	4.17
Uttar Dinajpur	0.95	1.01
Sector:		
Chemical products (incl. pharma)	8.41	7.88
Construction materials	14.14	13.97
Electrical & electronic goods	12.30	13.63
Food, drink & tobacco	11.74	10.27
Household goods	2.68	2.69
Machines & equipment	14.88	15.92
Metals	7.84	8.80
Mining & energy	2.20	1.99
Other	6.62	6.57
Rubber & plastic	3.53	3.76
Textiles	8.70	7.60
Wood & paper	6.96	6.92
Observations	177,973	106,775

This table presents the distribution of firms in percentage by size categories, districts, and sectors for both the universe of firms in the VAT data ('VAT data') and the sample of firms after assigning castes based on surnames and pruning the network, as described in the text ('Cleaned sample').

Table A2: Caste effect on firm exit

	(1) Exit	(2) Exit	(3) Exit	(4) Exit
Own caste share, demeaned (volume)	0.142*** (0.022)	0.123*** (0.021)	0.124*** (0.026)	0.202*** (0.035)
Age	-0.091*** (0.002)	-0.089*** (0.002)	-0.087*** (0.003)	-0.085*** (0.004)
Log sales	-0.263*** (0.006)	-0.284*** (0.005)	-0.342*** (0.007)	
Observations	227,716	227,716	227,716	227,716

This table presents the untransformed coefficients from our Cox model estimating the probability of firm exit, as described in the text. The coefficient 'own-caste share' is the share of sales going to firms from the supplier's caste in a given year, compared to the industry average. Firm age is the time in years since registration. Stratification lists the variables that define the subgroups for comparison, i.e. subgroup-specific baseline. Standard errors are clustered at the Postcode X good level. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table A3: Robustness, Extensive margin

	2011 - 2016							2013		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$	$\mathbb{1}(Trade)$
Same caste	0.990***	1.004***	1.046***	0.936***	1.355***	1.050***	0.990***	1.003***	1.050***	1.420***
	(0.021)	(0.025)	(0.024)	(0.025)	(0.031)	(0.022)	(0.110)	(0.025)	(0.027)	(0.036)
Robustness	Main	4-digit NIC	2nd caste	West Bengali surnames	No large castes	No small castes	Product & Post Code clustered SEs	Main	All potential sellers	Recipe
Obs. (thousand)	202,496	116,756	202,496	96,664	119,417	175,288	202,496	33,665	133,804	169,859

This table presents results obtained from running specification (1), described in the text, on different samples of our data as described below. The variable 'same caste' is an indicator equal to 1 divided by the mean probability of trade in the respective sample if the two firms are in the same caste, 0 otherwise. Column (1) presents the results from our main specification as shown in Table 3. Column (2) restricts the sample to firms whose main good is defined at the 4-digit level. Column (3) uses the secondary caste, if available, to construct the 'Same caste' indicator. Column (4) restricts the sample to firms which have a name that explicitly connected to West Bengal in the original source. Column (5) excludes all firms from one of the 3 largest castes (Aguri, Baidya, Marwari). Column (6) excludes who have no potential trading partner from their own caste in the sample. Column (7) is based on the main sample, but uses two way-clustered standard errors by Client location and industry.
Columns (8) - (10) use only data from the year 2013: Column (8) uses the main sample described in the text. Column (9) uses the same procedure to identify potential suppliers, but uses the full set of potential trading partners. Column (10) uses recipes to construct the set of potential suppliers as describe in the text, taking a 25% subset of potential trading partners. All columns include fixed effects for each interaction of the location (postcode) of the supplier and the location (postcode) of the client, supplier $\times$ year and client $\times$ year fixed effects and fixed effects for each interaction of the product sold by the supplier and the product sold by the client. If not indicated otherwise, standard errors clustered two-way at the level of the supplier and the client in parentheses. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table A4: Robustness, Intensive margin

	(1) Log. trade	(2) Log. trade	(3) Log. trade	(4) Log. trade	(5) Log. trade	(6) Log. trade	(7) Log. trade
Same caste	0.183***	0.195***	0.188***	0.139***	0.248***	0.178***	0.183***
	(0.009)	(0.011)	(0.009)	(0.012)	(0.013)	(0.009)	(0.023)
Robustness	Main	4-digit NIC	2nd caste	West Bengali castes	No large castes	No small castes	Product & Post Code clustered SEs
Obs. (thousand)	1,461	885	1,461	620	866	1,196	1,461

This table presents results obtained from running specification (1), described in the text, on different samples of our data as described below. The variable 'same caste' is an indicator equal to 1 if the two firms are in the same caste, 0 otherwise. Column (1) presents the results from our main specification as shown in Table 3. Column (2) restricts the sample to firms whose main good is defined at the 4-digit level. Column (3) uses the secondary caste, if available, to construct the 'Same caste' indicator. Column (4) restricts the sample to firms which have a name that explicitly connected to West Bengal in the original source. Column (5) excludes all firms from one of the 3 largest castes (Aguri, Baidya, Marwari). Column (6) excludes who have no potential trading partner from their own caste in the sample. Column (7) is based on the main sample, but uses two way-clustered standard errors by Client location and industry. All columns include fixed effects for each interaction of the location (postcode) of the supplier and the location (postcode) of the client, supplier $\times$ year and client $\times$ year fixed effects and fixed effects for each interaction of the product sold by the supplier and the product sold by the client. If not indicated otherwise, standard errors clustered two-way at the level of the supplier and the client in parentheses. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table A5: Caste effect and firm age

	(1) $\mathbb{1}(Trade)$	(2) $\mathbb{1}(Trade)$	(3) $\mathbb{1}(Trade)$	(4) Log. trade	(5) Log. trade	(6) Log. trade
Same caste	0.990***	0.978***	0.960***	0.183***	0.184***	0.185***
	(0.021)	(0.021)	(0.022)	(0.010)	(0.010)	(0.010)
Same caste × Experienced client		0.028			-0.002	
		(0.022)			(0.012)	
Same caste × Experienced supplier			0.069**			-0.003
			(0.027)			(0.013)
Obs. (thousand)	202,496	202,496	202,496	1,461	1,461	1,461

This table presents results obtained from running specification (1) augmented with an interaction term. The variable 'same caste' is an indicator equal to 1 if the two firms are in the same caste, 0 otherwise. The variables 'experienced supplier' and 'experienced client' are indicators whether the (potential) supplier, respectively client, were registered in 2005 or earlier. In columns (1), (2), (3), the variables 'same caste' and all of its interaction terms are divided by the mean probability of trade in the sample (.00726) All columns include fixed effects for each interaction of the location (postcode) of the supplier and the location (postcode) of the client, supplier × year and client × year fixed effects and fixed effects for each interaction of the product sold by the supplier and the product sold by the client. Standard errors clustered two-way at the level of the supplier and the client in parentheses. Significance levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table A6: Varna and caste

	(1) $\mathbb{1}(Trade)$	(2) $\mathbb{1}(Trade)$	(3) $\mathbb{1}(Trade)$	(4) Log. trade	(5) Log. trade	(6) Log. trade
Same castegroup	0.126***		0.064***	0.037***		0.014***
	(0.005)		(0.005)	(0.005)		(0.005)
Same caste		0.990***	0.967***		0.183***	0.179***
		(0.021)	(0.021)		(0.009)	(0.009)
Obs. (thousand)	202,496	202,496	202,496	1,461	1,461	1,461

This table presents results on larger caste groups' (varnas) effect on trade, running a regression based on our main specification (1). The variable 'same caste' is an indicator equal to 1 if the two firms are in the same caste, 0 otherwise; the variable 'same caste group' is one if the two firms belong to the same caste group (SC, ST, OBC, non-scheduled) as outlined by the "West Bengal Scheduled Castes, Scheduled Tribes and Other Backward Classes Development Finance Corporation". In columns (1), (2), (3), the variables 'same caste' and 'same caste group' are divided by the mean probability of trade in the sample (.00726) Standard errors clustered two-way at the level of the supplier and the client in parentheses. Significance levels: *p<0.1; **p<0.05; ***p<0.01.

Table A7: Externally calibrated parameters

Parameter	Definition	Value	Source
α	Labor cost share	0.4	Boehm and Oberfield (2020)
μ	Markup	1.34	De Loecker et al. (2016)
X	Aggregate final demand	1	Normalization
σ_ε	Pair matching cost dispersion	4	Bernard et al. (2022)
mean $\mathcal{C}$	Share of same-caste pairs	3.4%	Calibrated