Cultural Proximity and Inter-firm Trade*

Brian C. Fujiy† U.S. Census Bureau

Gaurav Khanna‡ UCSD

Hiroshi Toma§ Texas A&M University

First draft: April, 2022
This draft: August, 2025

Abstract

Low-quality institutions are prevalent in emerging economies, generating micro-level trade frictions. Consequently, firms may rely on informal institutions and socio-cultural ties between firm owners, to overcome such frictions. Using new microdata on firm-to-firm trade from India with information on prices, transactions, and caste and religious connections, we find that higher caste and religious proximity reduces prices, and fosters trade at intensive and extensive margins. We provide supporting evidence on such proximity alleviating contracting frictions. We formalize these findings in a quantitative inter-firm trade model with cultural proximity between firm owners. A policy counterfactual analysis indicates that an economy composed of culturally closer firms features lower costs, lower prices, higher sales, and higher welfare than an economy with culturally distant firms.

Keywords: cultural proximity, inter-firm trade, domestic trade, contracting frictions, contract enforcement, social inclusion

JEL Codes: D51, F19, O17

*We thank Treb Allen, Miriam Artiles, Mathias Buehler, Paula Bustos, Vasco Carvalho, Thomas Chaney, Ornella Darova, Kevin Donovan, Ben Faber, Farid Farrokhi, Alan Griffith, Amit Khandelwal, Andrei Levchenko, Kanika Mahajan, Rocco Macchiavello, Alexey Makarin, Yuhei Miyauchi, Joan Monras, Swapnika Rachapalli, Kim Ruhl, Jagadeesh Sivadasan, Sebastian Sotelo, Tillmann von Carnap, Daniel Xu, and Yingyan Zhao for insightful comments. We thank seminar participants at the University of Michigan, BSE Summer Forum, JADE-CEPR-TIME Conference, LMU Economics of Firms and Labor Workshop, MWIEDC Conference, NEUDC Conference, PacDev Conference, STEG Theme Workshop, WAITS Conference, and UDEP Workshop for Young Economists for their useful suggestions. We thank Manaswini Bhalla, Manisha Goel, and Manaswini Rao for sharing their datasets. Eli Mogel and Tyler Spencer provided excellent research assistance. We also thank the 2021 Hellman Fellowship for their generous support for this project. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau.

†*Email:* brian.c.fujiy@census.gov

‡Corresponding author. *Email:* gakhanna@ucsd.edu. *Address:* School of Global Policy and Strategy, University of California San Diego, 9500 Gilman Drive, La Jolla, , CA

§*Email:* htoma@tamu.edu

1 Introduction

Non-economic forces, such as religion, language, caste, and community ties, may drive economic outcomes. The role of socio-cultural factors on networks and agent behavior has been well documented in entrepreneurship, loan access, labor markets, marriage, development, and cross-country international trade (Artiles, 2023; Fisman et al., 2017; Goraya, 2023; Munshi and Rosenzweig, 2016; Rauch, 1996; Rauch and Trindade, 2002; Startz, 2016). Nevertheless, the mechanisms by which such proximity shapes trade between firms remain less understood. Understanding how and why socio-cultural linkages affect inter-firm trade, potentially allows policy-makers to better leverage social inclusion programs and foster economic development.

In this paper, we define cultural proximity to be the links between firm owners on caste and religious grounds. We provide empirical evidence on the role of such cultural proximity in inter-firm trade. To do this, we leverage a unique dataset of transactions between firms from a large Indian state, along with data on firm owners' names and their cultural proximity derived from India's caste and religious system. We report three new stylized facts. First, culturally closer firms report higher sales between them: the higher the cultural proximity, the higher the trade on the intensive margin. Second, culturally closer firms are more likely to trade with each other. That is, the higher the cultural proximity, the higher the trade on the extensive margin as well. Third, culturally closer firms report lower unit prices in their transactions. All these results are robust to an array of high-dimensional fixed effects, including seller and buyer fixed effects, origin-by-destination fixed effects (and for specifications with product and time, seller-by-product, and product-by-month fixed effects).

We then turn to explore various mechanisms, and find evidence most in line with the importance of contract enforcement. First, we show that the effect we find of cultural proximity on trade is driven by differentiated products, which often rely on either formal or informal contract enforcement (Nunn, 2007; Rauch, 1999).[1] Second, cultural proximity matters the most when institutional quality, proxied by the court quality in the districts where the trade partners belong, is particularly low. We argue that, in a setting with low institutional quality and poor contract enforcement, firms that trade differentiated goods rely on informal institutions (i.e., cultural proximity) as a substitute for the imperfect formal ones. We understand these findings as evidence that cultural proximity may relate to contract enforcement and trust (Munshi, 2019, 2014).[2]

[1] Differentiated goods do not trade in exchanges and are not homogeneous, but are branded and specific to certain producing firms. In a country with market imperfections as India, firms can easily renege on their commitments. Suppliers and buyers in differentiated goods markets are not easily replaceable. In such cases, trade will increase when firms trust and know each other, that is, when they are culturally close.

[2] Munshi (2019) uses survey data to show that Indians trust people from their caste.

We further find that the more varieties a firm sells or buys, the more the trade intensity is affected by cultural proximity. We posit that the larger the number of different varieties a firm sells or buys, the more firms it has to negotiate with, which increases the contracting frictions it faces. Then, to minimize the contracting frictions they face, firms will rely more on trading with culturally closer firms they trust. Moreover, our data is special because it records which transactions are canceled. We find evidence that shows it is less likely that there will be canceled transactions when firms are culturally closer, which highlights how cultural proximity relates to contract enforcement.

We do not find sufficient evidence that hierarchies (and preference-based discrimination) across social groups matter, or that linguistic distance and the specialization in certain goods matter for our cultural proximity results. To analyze whether our results are caused by vertical social hierarchies and discrimination across cultural groups, we study asymmetric effects in those transactions where one firm is placed higher than the other based on the caste-based hierarchy, allowing us to test for preference-based discrimination across the social hierarchy. In other tests, we find our results are less likely to be driven by firms sharing the same language or sharing specialization in the production of certain goods. Also, we find little evidence that caste-based cartels are significant in this context. However, family links, based on firm owners having the same surname, may be important.

We then conduct a counterfactual analysis to examine the significance of cultural proximity in inter-firm trade. We build a quantitative trade model with cultural proximity between firms. Firms optimally decide whom to trade with, subject to matching fixed costs, and how much to trade with, subject to iceberg trade costs. In line with our empirical findings, we allow these costs to depend (positively or negatively) on how culturally close firms are.[3]

Our matching function generates the same primary independent variable of interest in our reduced-form section, and the model derives equations that precisely match their empirical counterparts in the previous section. We use these equations to estimate the key parameters of the model: the semi-elasticity of the trade cost to cultural proximity and the semi-elasticity of the matching cost to cultural proximity. Our model allows us to estimate both of these parameters externally. In line with our stylized facts, we find a negative semi-elasticity of both the intensive and extensive margin of trade to cultural proximity. This implies that the closer two firms are in cultural terms, the lower the trade and matching costs are. Therefore,

[3] On the extensive margin, cultural proximity between firms can reduce matching costs since culture encodes useful information for sellers to decide who to sell to (Ali and Miller, 2016; Balmaceda and Escobar, 2017). On the intensive margin, sellers charge a premium to buyers arising from contracting frictions. Given the risk of reneging on the contract or delaying payment, sellers determine the charged premium to buyers, and so affect the intensive margin of trade (Boehm and Oberfield, 2020).

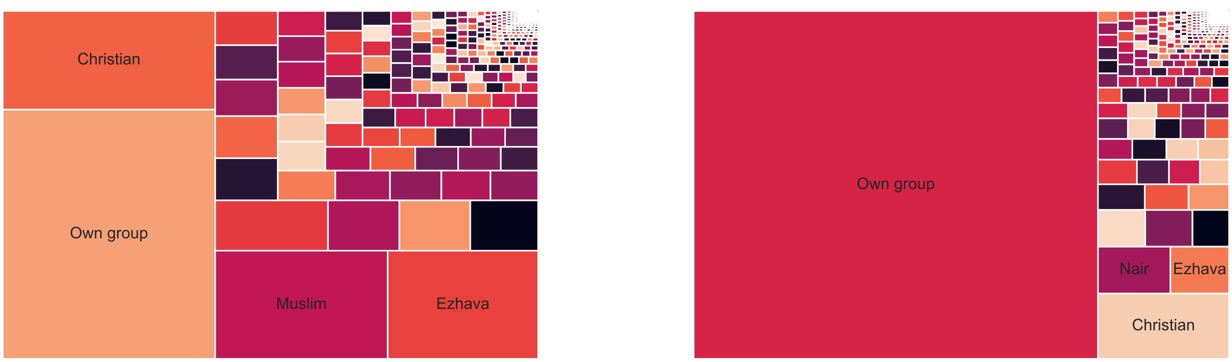
the higher the cultural proximity for a pair of firms, the higher the trade is on both the intensive and extensive margins, and the lower the prices charged.

We then use the model and estimated parameters to quantify the implications for welfare and other aggregate outcomes of implementing different policies. First, we evaluate the effects of social mixing/inclusion (i.e., firms become culturally the closest possible) and social isolation policies (i.e., firms become culturally the furthest possible). Second, we study the effects of a policy that reduces contracting frictions, such that firms rely less on cultural proximity when trading (i.e., trade and matching costs become less sensitive to cultural proximity). We find that welfare increases by 1.90 percent under a diversity-friendly social inclusion policy. In contrast, welfare falls by 0.93 percent when we evaluate the effects of social isolation or exclusion. Finally, we show that policies that reduce contracting frictions raise welfare by 0.95 percent by reducing the reliance of trade on cultural links.

Figure 1: Probability-weighted sales decomposition of largest cultural groups

(a) Largest Hindu group: Nair

(b) Largest non-Hindu group: Muslims



Notes: The figure shows the decomposition across buyers for the largest Hindu and non-Hindu cultural groups measured by probability-weighted sales. For instance, the Nair and Muslims accounted for 5.39 and 19.68 percent of total probability-weighted sales, respectively. Data corresponds to the year 2019.

The analysis of cultural proximity is especially relevant for developing countries, where agents face several contracting frictions and, consequently, rely more on non-economic forces. In particular, India has a society that follows the parameters of a caste system, which also intertwines with the different religious groups.[4] In this case, cultural proximity naturally arises as a product of the inherent hierarchical structure of the caste system and the different

[4]In this paper, we consider the caste system and religious groups as a proxy for cultural groups. There is a large historical legacy for the caste system as a discriminatory device, which we consider. Even though there is an active agenda of the government to implement policies that hinder caste-based discrimination, it is still used by Indians as a way to determine how similar individuals are between them.

religions. Related to this, Figure 1 shows an example of how trade between cultural groups occurs, in a selected subset of our data. We can see that there are cultural groups that are bound to trade more or less with other cultural groups. We thus ask whether cultural proximity, measured as the cultural group-based distance between firms, can determine trade.

While we use the term 'cultural proximity,' our work is not about how cultural norms affect economic outcomes in India (Bau et al., 2023). Rather than studying the direct impacts of cultural norms, we are studying the consequences of links and ties along religious and caste lines. Indeed, we do not use information about the norms and behaviors of a specific cultural group in determining economic outcomes. Instead, our work emphasizes the importance of bilateral links along caste, community, and religious lines, without relying on any information about the practices of any particular group.

Furthermore, we do not expect names to affect firm behavior directly. Rather, we make the case that firms are human entities, and are run by people who make contracts with other people. This is especially true in our context, where we show that most firms are small, rather than large conglomerates. In entering into such contracts, individual firm owners need to build trust and ensure these contracts be enforced. In such a setting, caste and religious ties influence contracting behavior and, consequently, firm trade.

We contribute to two strands of the literature. First, we speak to the role of cultural proximity on trade (Bandyopadhyay et al., 2008; Boken et al., 2022; Guiso et al., 2009; Macchiavello and Morjaria, 2015; Rauch, 1996; Rauch and Casella, 2003; Rauch and Trindade, 2002; Richman, 2006; Schoar et al., 2008; Startz, 2016; Zhou, 1996). Cultural differences in international trade or across administrative regions also interplay with non-cultural barriers. In contrast, we use transaction-level firm-to-firm data to explore the effect of culture on trade that does not rely on cross-border variation. This allows us to isolate the mechanisms at a granular level, and to provide a framework that quantifies the role of cultural proximity for inter-firm trade.[5] We also connect to the importance of cultural or social proximity on other economic outcomes, like entrepreneurship (Goraya, 2023), finance (Fisman et al., 2017), the composition of the board of directors (Faia et al., 2021), and labor markets (Munshi and Rosenzweig, 2016). We complement this work by examining how culture affects

[5] In ongoing work, Boken et al. (2022) also shows the role of cultural proximity for inter-firm trade. We mainly distinguish ourselves by leveraging data on prices, which allows us to estimate how cultural proximity influences inter-firm trade through the alleviation of contracting frictions. Additionally, we explore a rich set of mechanisms, as our data contains information to test how cultural proximity matters for inter-firm trade through discrimination (caste hierarchies), order cancellations, firm survival over age, and how cultural proximity matters for firm-level complexity. Leveraging other data, we implement a Heckman selection bias correction model to estimate trade elasticities following Helpman et al. (2008).

trade frictions. Understanding the source of these frictions or even leveraging them is key to adequately determining the consequences of economic policy.

Second, we contribute to the literature on social cohesion (Alan et al., 2021; Alesina and Giuliano, 2015; Alesina et al., 2021; Alesina and Reich, 2015; Bazzi et al., 2019; Depetris-Chauvin et al., 2020; Gradstein and Justman, 2019; Ritzen et al., 2000) and contract enforcement policies.[6] We contribute by implementing a policy counterfactual analysis that quantifies the aggregate effects of education and court quality policies through cultural proximity.

The rest of the paper is structured as follows. In Section 2 we provide a brief review of the caste system in India, describe our new datasets, and explain how we construct firm-level trade and cultural proximity variables. In Section 3, we report our stylized facts, study the mechanisms driving the effect of cultural proximity on trade, and robustness checks. In Section 4, we briefly describe the model, how we estimate the model, and perform a counterfactual analysis. Section 5 concludes.

2 Background, data and construction of variables

2.1 Caste and Religion in India

Indian society is heavily influenced by the parameters of a caste system: a hierarchical system that has prevailed in the country since around 1,500 BC and that still rules its economy. People are classified across four possible groups called *Varnas*, in a hierarchical order, with an additional group *Dalits* positioned below. Varnas are further comprised of sub-groups called *Jatis*, determined by factors such as occupation, geography, tribes, or language. We use Jatis as an appropriate classification for studying economic networks (Munshi, 2019).

In our definition of cultural groups, we also consider religious groups. The caste system was originally based on the Hindu religion. While other religions in India did not historically follow the caste system, they do relate to it today: the other non-Hindu religions work as cultural groups of their own. We leverage information on firm owners belonging to both caste and non-Hindu religious groups to construct our measure of cultural proximity.

Importantly, caste and religion have been shown to play a prominent role in the Indian economy. For instance, caste and religion strongly determine access to credit, and as a

[6] Contracting frictions can be either formal or informal. We show that informal channels, such as cultural proximity, matter in the aggregate when implementing policies.

result, entrepreneurship and firm growth (Goraya, 2023). Indeed, credit access, which is crucial for firm production, is meaningfully mediated by caste and religious groups (Fisman et al., 2017). Given the importance of group-based networks in entrepreneurial activity, caste and religious networks are also influenced by political linkages, whereby entrepreneurs from the same group benefit from local political representation (Bhalla et al., 2024). Outside of firms, individual-level networks in the economy rely on caste linkages for the sharing of information, risk, and insurance (Munshi and Rosenzweig, 2016). More broadly, caste does affect firm and individual networks, and the concentration of certain groups in certain types of products (Munshi, 2019).

2.2 Data

Firm-to-firm trade. We obtain a new firm-to-firm trade dataset for a large Indian state provided by the state's corresponding tax authority.[7] We use daily transaction-level data from January 2019 to December 2019, as long as at least one node of the transaction (either origin or destination) was in the state. This data exists due to the creation of the E-Way bill system in India in April 2018, where firms register the movements of goods online for tax purposes. This is a major advantage over traditional datasets collected for tax purposes in developing countries, since the E-Way bill system was created to increase tax compliance significantly.[8]

The state has a diversified production structure, with roughly 50 percent urbanization rates, and high levels of population density. To compare its size in terms of standard firm-to-firm transaction datasets, the population of this Indian state is roughly three times the population of Belgium, seven times the population of Costa Rica, and double the population of Chile. In addition, we can uniquely measure product-specific prices for each transaction, along with the usual measures of the total value traded.

Each transaction reports a unique tax code identifier for both the seller and buyer. We use these identifiers to merge this data with other firm-level datasets. We also have information

[7] While we use the term 'firm' in most parts of the paper, these data are actually at the more granular establishment level. The name of the Indian state must remain confidential as per our signed Memorandum of Understanding with the relevant tax authority.

[8] Tax evasion rates are thought to have fallen with the E-Way bill system, given how it is implemented. A selling establishment must online register the transaction, and print out a receipt that the driver of the transportation (usually a truck) must carry with them while transporting the product. If the driver is stopped or checked at any of the numerous checkpoints, and fails to produce a receipt, the goods are confiscated. Furthermore, the earlier VAT regime was only for large firms, and the new GST system was aimed at including smaller firms too. For more details about the new E-Way bill system, see https://docs.ewaybillgst.gov.in/

on all the items contained within the transaction, the value of the transaction, the 6-digit HS code of the traded items, the quantity of each item, and the units the quantity is measured in. Since the data report both value and quantity of traded items, we construct unit values for each transaction. Each transaction also reports the pincode (zip code) location of both selling and buying firms. By law, any person dealing with the supply of goods and services whose transaction value exceeds 50,000 Rs (600 USD) must generate E-way bills. Transactions that have values lower than 600 USD can also be registered, but it is not mandatory. There are three types of recorded transactions: (i) within-state trade, (ii) across-state trade, and (iii) international trade. For this paper, we ignore international trade.

Firm owner names. Information about the names of firm owners comes from two different sources. The first is also provided by the tax authority of the Indian state, which is a set of firm-level characteristics for firms registered within our large state. Among these variables, we are provided with the name of the owner, directors and/or representatives of the firm.

To obtain firm-level characteristics of firms not registered in this state, we scrape the website *IndiaMART*,[9] the largest e-commerce platform for business-to-business (B2B) transactions in India. The website comprises firms of all sizes. By 2019, the website registered around 5-6 million sellers scattered all around India. Most importantly, this platform provides the name of the owner of the firm and the unique tax code identifier. Thus, we use the platform to obtain these variables for out-of-state firms.

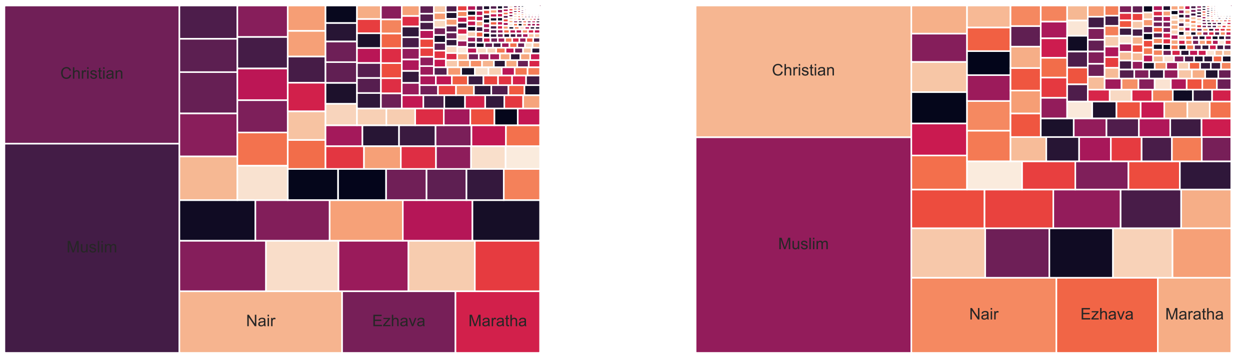
Matching owner names to cultural groups. We follow Bhagavatula et al. (2018) to match owner names to their Jatis (if the owners are of the Hindu religion) or to their religion (in case the owners are not Hindu). Their procedure consists of using scraped data from Indian matrimonial websites that contain information on names, castes, and religions. They train a sorting algorithm that uses names as inputs and gives a probability distribution across cultural groups per name as outputs. We match these probability distributions to each owner's name in our dataset. Notice that our notion of cultural group belonging is probabilistic and not deterministic. This probabilistic approach is more relevant to our setup since, when firm owners trade with each other, they do not know each other's cultural group *ex ante*. Our sample finally consists of 452 cultural groups.

Merged dataset. For the analytical part, we merge the three previous datasets. We end up with a sample that contains information from 21,497 unique firms for the year 2019, of which there are 10,010 sellers and 16,403 buyers. We drop any registered transaction in which the seller and the buyer are the same firm. In total, the dataset comprises approximately 800

[9] https://www.indiamart.com/

Figure 2: Probability-weighted sales and purchases across cultural groups

(a) Sales (b) Purchases



Notes: Figure shows the decomposition of the probability-weighted sales and purchases across the 452 cultural groups in our dataset. The size of the rectangles reflects the share of sales and purchases. Data corresponds to the year 2019.

thousand transactions, equivalent to 1.2 billion US dollars. Each firm is linked to a unique pincode. Finally, we assign a sector to each firm based on the ISIC codes of the goods sold. We provide more aggregate and firm-level statistics in the Appendix Table A1.

There is a large degree of heterogeneity across firms and cultural groups. Appendix Figure A1 shows that there is a significant dispersion in sales per firm, although most of the firms within the dataset can be classified as small. Furthermore, to provide a summary of the heterogeneity of cultural groups present in the firm-to-firm trade data, we show the distribution of probability-weighted sales and purchases across cultural groups in Figure 2.

2.3 Construction of variables

Inter-firm trade variables. The firm-to-firm dataset provides information at the transaction level between any two registered firms. More specifically, we have information on (i) transaction-level unique identifiers, (ii) seller and buyer unique identifiers, (iii) the 6-digit HS description of the traded goods in each transaction, (iv) the total value of the transaction in rupees per type of good involved in each transaction, and (v) the number of units sold of each good in each transaction.

For every seller/buyer pair, we construct total sales, the total number of transactions, and unit values. For the total sales, we add up all the sales between each given pair of firms in our sample. We do the same with the total number of transactions. To obtain prices, we calculate the unit values. We first calculate the total amount sold and the total units sold of each good at the 6-digit HS level between each given pair of firms in our sample. Then, we divide the total amount sold by the number of units sold for each good. All of these variables use data for the year 2019 only.

Cultural proximity. Consider the set $\mathcal{X}$ of cultural groups, where $|\mathcal{X}| = X = 452$ in our final dataset. Since not all names are deterministically matched to a cultural group, each firm in our dataset has a discrete probability distribution over the set X of cultural groups. In particular, every firm ν has a probability distribution $\boldsymbol{\rho_\nu} = [\rho_\nu(1), \ldots, \rho_\nu(X)]$, such that $\sum_{x=1}^{X} \rho_\nu(x) = 1$. In this part, we distinguish between the probability distribution over cultural groups of the seller and the probability distribution over cultural groups of the buyer. Define $\rho_\nu(x)$ as the probability of seller ν of belonging to cultural group x. Similarly, define $\rho_\omega(x)$ as the probability of buyer ω of belonging to the cultural group x. Based on these two distributions, we construct the following measure of cultural proximity: the Bhattacharyya coefficient (Bhattacharyya, 1943).

The Bhattacharyya coefficient between seller ν and buyer ω measures the level of overlap between two different probability distributions.[10] We define it as

$$BC(\nu, \omega) = \sum_{x=1}^{X} \sqrt{\rho_\nu(x)\, \rho_\omega(x)}.$$

Because $0 \leq \rho_\nu(x) \leq 1$ and $0 \leq \rho_\omega(x) \leq 1$, we have that $0 \leq BC(\nu, \omega) \leq 1$. On the one hand, $BC(\nu, \omega) = 0$ means the seller has a completely different probability distribution from the buyer's. In our context, this means the seller and the buyer have almost no chance of belonging to the same cultural group or that their cultural proximity is the farthest. On the other hand, $BC(\nu, \omega) = 1$ means the seller has the same probability distribution as the buyer. This implies that the seller has the same probability of belonging to a group of certain cultural groups as the buyer or that their cultural proximity is the closest possible.[11]

[10] Notice the Bhattacharyya coefficient is not the Bhattacharyya distance, which is defined as $BD(s, b) = -\log(BC(s, b))$. We prefer the Bhattacharyya coefficient because it is easier to interpret.

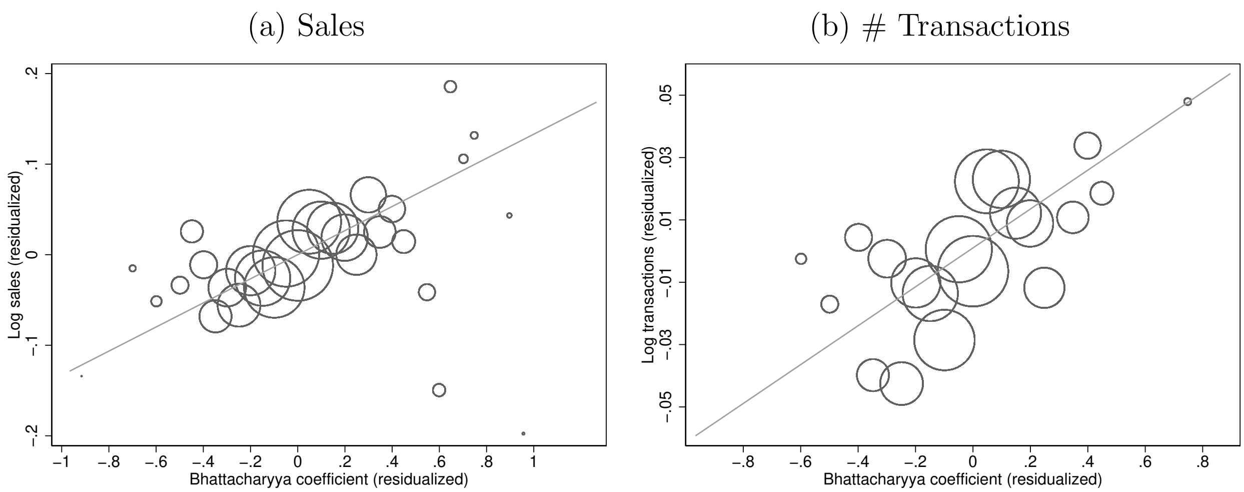
[11] For our purposes, it is important that the cultural proximity measure we use is symmetric. To see why, consider an example where, in our dataset, we have a transaction between a seller ν and a buyer ω, from which we obtain $BC(\nu, \omega)$. Further, assume that in our dataset we record a second transaction in which the roles of the firms revert (i.e., the buyer becomes the seller and vice versa), so we calculate $BC(\omega, \nu)$. Regardless of the roles the firms take in this second transaction, we want their cultural proximity to remain constant, as the membership of cultural groups is fixed. This goal is achieved through the means of a symmetric proximity measure, and the Bhattacharyya coefficient complies with this symmetry requirement, as $BC(\nu, \omega) = BC(\omega, \nu)$.

In robustness checks, we use the Kullback and Leibler (1951) divergence as an alternative measure of cultural proximity (Appendix B.2). In this case, all our results are qualitatively similar, and remain statistically significant.

3 Stylized facts

Fact 1: Cultural proximity fosters trade. We first discuss results related to the intensive margin of the firm-to-firm trade. Figure 3 shows the residualized scatterplots between the Bhattacharyya coefficient and two intensive margin measures: total sales between two firms and total transactions between two firms. Both measures capture the totality of pairwise sales and transactions for the firms in the dataset in the year 2019. The scatterplots show that a higher Bhattacharyya coefficient (buyer and seller are probabilistically more alike in their cultural group) is related to a higher amount of sales and transactions.

Figure 3: Effect of cultural proximity on trade, intensive margin



Notes: Results are residualized of seller fixed effects, buyer fixed effects, and log distance. Equally distanced bins formed over the horizontal axis. The size of the bubbles represents the number of transactions in each bin. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. Total sales and total number of transactions represent the totality of pairwise sales and transactions of the firms in the dataset. Data corresponds to the year 2019.

We confirm the findings using a gravity equation. For transactions from firm ν to firm ω in our sample we estimate

$$\ln y\left(\nu,\omega\right) = \iota_{\nu} + \iota_{\omega} + \delta BC\left(\nu,\omega\right) + \eta\Xi\left(\nu,\omega\right) + \varepsilon\left(\nu,\omega\right), \tag{1}$$

where $y(\nu,\omega)$ is either the total sales $n(\nu,\omega)$ or total transactions $t(\nu,\omega)$ from seller ν to buyer ω, $BC(\nu,\omega)$ is the Bhattacharyya coefficient, and ι_ν and ι_ω are seller and buyer fixed effects, respectively. $\Xi(\nu,\omega)$ denotes distance control variables—i.e. the Euclidean distance between pincodes in which the firms are located $dist(\nu,\omega)$ or a district-level origin-destination fixed effect ι_{od}), and a distance measure based on name similarity $dist^{name}(\nu,\omega) = \frac{L(\nu,\omega)}{max(len(\nu),len(\omega))}$, where the numerator $L(\nu,\omega)$ is the Levenshtein distance in names, and the denominator is the maximum in between the length of the last name of seller ν and buyer ω.[12] As mentioned before, our measures of total sales and total transactions capture the sum of all the pairwise sales and transactions of the firms in the dataset in the year 2019.

Columns 1 to 4 in Table 1 present the results of the intensive margin estimation, which confirm the preliminary findings from Figure 3. Columns 1 and 2 show that, on average, there will be a higher amount of sales and transactions between a pair of firms when these firms are more alike in cultural terms. Columns 3 and 4 show that these results remain strong after including origin-destination fixed effects, which account for geographic distance but also control for other features that might arise between a pair of locations, such as different terrains, different languages, location-specific cultural ties, historical ties, etc.

Columns 5 to 8 in Table 1 also consider the role of name proximity between firm owners. Results show that names do matter: the closer the names are, the more the trade. Yet, even after controlling for name distance, the result for cultural proximity and trade remains almost identical. That is, even after considering how similar the names are, cultural proximity facilitates more trade.

The results suggest that the cultural proximity measure $BC(\omega,\nu)$ contains more information than just people sharing a name or having family ties. For instance, assume two firm owners belong to the same caste (or even the same family) but have different last names. Our cultural proximity measure captures the fact that these two people from the same caste but with different names are culturally close, and this cultural proximity matters over and above any proximity in names alone.

Fact 2: Cultural proximity increases the likelihood of trading. Next, we estimate the extensive margin relationship. Given the size of our full dataset, the number of potential extensive margin links is computationally large. For tractability, we modify our sample.

[12]The Levenshtein distance measures how many changes have to be made to the seller's last name in order to match the buyer's last name (i.e., changing a character, erasing a character, or adding a character). Moreover, we divide this distance measure to normalize it to the number of characters contained in the last names of the seller and buyer. Thus, the larger the distance measure between names $dist^{name}$, the more different the last names of the seller and buyer are.

Table 1: Effect of cultural proximity on trade, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.100***	0.066**	0.129***	0.076***
	(0.033)	(0.027)	(0.034)	(0.028)
Log dist.	-0.023	-0.065***		
	(0.015)	(0.011)		
Obs.	32,678	32,678	32,843	32,843
Adj. R2	0.415	0.359	0.410	0.356
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.
	(5)	(6)	(7)	(8)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.099**	0.067**	0.129***	0.083***
	(0.039)	(0.031)	(0.039)	(0.031)
Log dist.	-0.020	-0.062***		
	(0.017)	(0.013)		
$dist^{name}$	-0.378***	-0.303***	-0.415***	-0.290***
	(0.094)	(0.075)	(0.093)	(0.073)
Obs.	24,057	24,057	23,999	23,999
Adj. R2	0.412	0.364	0.404	0.362
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.

Notes: Columns show the results of estimating Equation 1. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer in the year 2019, and it is equal to 0 otherwise. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The larger the name distance $dist^{name}$, the more different the names of the firm owners are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

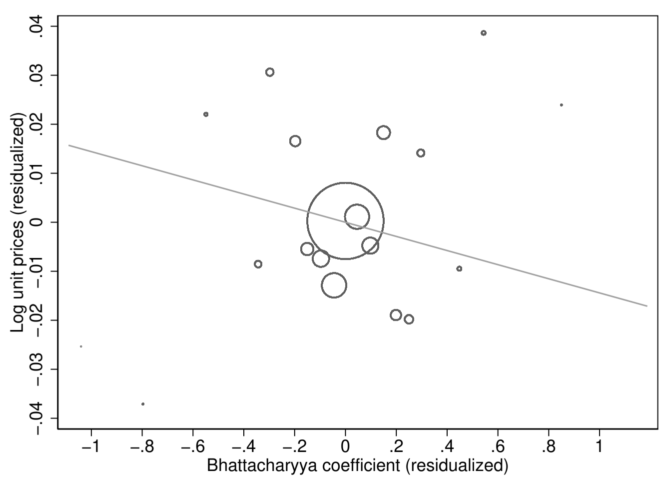
In the first place, we construct a sample with all possible combinations of in-state buyers and in-state sellers with cultural group information. Then, we proceed to drop all potential transactions that include unfeasible sectoral combinations. That is, we drop the combinations of firms that are involved in productive sectors that never recorded a transaction in the data. Finally, we drop all unfeasible transactions based on distance. This is to say, we drop the combinations of firms where the seller is further away than the maximum recorded distance for the in-state buyer or vice versa.

With this sample, we construct a trade indicator variable $tr(\nu, \omega)$ which is equal to 1 if there is any kind of trade between firms ν and ω in the year 2019, and 0 otherwise. With this variable, we estimate a gravity-type specification:

$$tr(\nu, \omega) = \iota_\nu + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \varepsilon(\nu, \omega, t). \tag{2}$$

Columns 5 and 6 of Table 1 present the extensive margin results. We find that the higher the Bhattacharyya coefficient, the more likely it is that two given firms will trade.

Figure 4: Effect of cultural proximity on prices



Notes: Results are residualized for seller fixed effects and HS code fixed effects. Sectors are defined according to the 6-digit HS classification. Equally spaced bins formed over the horizontal axis. The size of the bubbles represents the number of transactions in each bin. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The unit prices represent the unit value of good g (at the 6-digit HS classification) sold by firm ν to firm ω in month t. Data corresponds to the year 2019.

Fact 3: Cultural proximity lowers prices. Figure 4 now uses buyer-seller-product-month groups and shows the residualized scatterplots between the similarity measure and the unit prices. We see that the higher the Bhattacharyya coefficient between two firms involved in a transaction, the lower the price that will be charged. To confirm the results, we work with transaction-level data and estimate:

$$\ln p_g(\nu, \omega, t) = \iota_{\nu g} + \iota_{gt} + \iota_\omega + \delta BC(\nu, \omega) + \eta \Xi(\nu, \omega) + \epsilon_g(\nu, \omega), \tag{3}$$

where $p_g(\nu, \omega, t)$ is the unit value of good g (at the 6-digit HS classification) sold by firm ν to firm ω in month t of year 2019, $\iota_{\nu g}$ is a seller-good fixed effect and ι_{gt} is a good-month

Table 2: Effect of cultural proximity on trade, extensive margin

	(1)	(2)
Dep. Variable	Trade Indicator	Trade Indicator
BC	0.0009***	0.0010***
	(0.0001)	(0.0001)
Log dist.	0.0001	
	(0.0000)	
Obs.	5,606,627	5,628,290
Adj. R2	0.617	0.0106
FE	Seller, buyer	Seller, buyer, origin×dest.

Notes: Columns show the results of estimating Equation 2. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer in the year 2019, and it is equal to 0 otherwise. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

fixed effect. We present the results in Table 3, which confirms the previous findings from Figure 4: the culturally closer firms are, the lower the unit value of the transactions.

3.1 Mechanisms

3.1.1 Contracting frictions

Differentiated goods and contract enforcement. To understand the underlying forces driving these empirical patterns, we explore the role of cultural proximity in alleviating contracting frictions. The lack of contract enforcement in developing countries inhibits trade as sellers or buyers may not comply with the terms of the contract. For instance, the buyer could hold up the seller by withholding payment after the buyer receives the shipped goods. Differentiated or relationship-specific goods are subject to more severe hold-up problems, as there are fewer alternative buyers and sellers of such products, and buyers may withhold payment knowing that these goods are not useful outside of the relationship. In that sense, differentiated goods rely on better contract enforcement (De Sousa et al., 2023).

Contract enforcement can be either formal (e.g., courts) or informal (e.g,. cultural proximity). Most of the literature has focused on the role of formal institutions in enforcing

Table 3: Effect of cultural proximity on prices

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices
BC	-0.069**	-0.069**	-0.066**	-0.045*	-0.040*	-0.039*
	(0.033)	(0.033)	(0.033)	(0.023)	(0.023)	(0.022)
Log dist.	0.023	0.023	0.028*			
	(0.016)	(0.016)	(0.017)			
Obs.	230,744	230,744	226,645	235,001	236,617	230,900
Adj. R2	0.932	0.932	0.935	0.933	0.925	0.936
FE	Seller×HS, buyer	Seller×HS, buyer, month	Seller×HS, buyer, month×HS	Seller×HS, buyer, origin×dest.	Seller×HS, buyer, month, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating Equation 3. Good g is defined according to 6-digit HS classification. Prices trimmed by 4-digit HS code at 5 and 95 percent. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Prices refer to unit prices of goods sold. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are multi-way clustered at the seller, 4-digit HS and origin-destination level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

contracts. We hypothesize and show evidence that cultural proximity alleviates contracting frictions for differentiated goods when formal contract enforcement is lacking (Nunn, 2007).

To bring in information about the type of product, we disaggregate our data at the transaction level. Then, we classify the goods into differentiated goods and non-differentiated goods based on on the classification developed by Rauch (1999).[13] We estimate the following specification:

$$\begin{aligned}\ln n_g(\nu,\omega,t) = \quad & \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu,\omega) + \xi\left(BC(\nu,\omega) \times \mathbb{I}_g^{diff}\right) \\ & + \eta \Xi(\nu,\omega) + \epsilon_g(\nu,\omega),\end{aligned} \tag{4}$$

where $n_g(\nu,\omega,t)$ are the sales going from firm ν to firm ω of good g in month t of year 2019, and $\mathbb{I}_g^{diff}$ is an indicator for differentiated goods.[14] Columns 2 and 3 of Table 4 present the

[13] According to Rauch (1999) differentiated goods are the goods not traded in organized exchanges or not reference priced in commercial listings. Differentiated goods have specific characteristics that "differentiate" (i.e., specialized goods, branded goods) them from other, more homogeneous types of goods. Because of their relative uniqueness in features, these goods are not as easily replaceable as non-differentiated goods and, as such, rely more on relationship-specific types of trade. This means sellers and buyers must face search frictions to match to a suitable trade partner and will likely not abandon the commercial matches they have already made.

[14] We use both the conservative and liberal classifications from Rauch (1999). The conservative classifi-

results. Our findings suggest that the baseline results of cultural proximity increasing trade are mostly driven by differentiated goods.

Table 4: Effect of cultural proximity on trade by types of good, intensive margin

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.099***	0.018	0.039	0.069**	-0.019	0.013
	(0.031)	(0.050)	(0.040)	(0.027)	(0.048)	(0.038)
$BC \times \mathbb{I}_g^{diff,con}$		0.122**			0.139**	
		(0.058)			(0.059)	
$BC \times \mathbb{I}_g^{diff,lib}$			0.097**			0.095**
			(0.047)			(0.047)
Obs.	174,352	174,352	174,352	177,584	177,584	177,584
Adj. R2	0.852	0.852	0.852	0.853	0.853	0.853
FE	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating Equation 4. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to the 6-digit HS classification. Sales were trimmed by 4-digit HS code at 5 and 95 percent. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and 4-digit HS level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the more culturally closer two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019). $\mathbb{I}_g^{diff,con}$ indicates the good g is a differentiated one according to the conservative classification of Rauch (1999). $\mathbb{I}_g^{diff,lib}$ indicates the good g is a differentiated one according to the liberal classification of Rauch (1999).

To understand the channel for why trade in differentiated goods depends on cultural proximity, we turn to the analysis of contract enforcement. We posit that, when facing poor contract enforcement and poor quality of institutions, firms trading differentiated goods must rely on alternative mechanisms that substitute the formal ones. Here, cultural proximity arises as a substitute for the trust and enforcement a well-functioning contract would have provided (Munshi, 2019, 2014).

To test this channel, we use data from Ash et al. (2021), and calculate two different measures of court quality between 2010 and 2018. First, for each court at the district level, we calculate the average number of days in between the filing of a case and its first hearing, $days^{first}$.

cation minimizes the number of goods classified as non-differentiated and, thus, has the largest amount of differentiated goods. The liberal classification maximizes the amount of goods classified as differentiated and has the largest number of differentiated goods.

Second, for each of these courts, we also calculate the average number of days between the filing and decision of a case, $days^{decision}$. The intuition is that the larger these two measures, the more congested and/or the worse contract enforcement is.

However, our data is at the pairwise level, so we need to construct a measure that captures the quality of the district court that corresponds to the seller on one side and to the buyer on the other side. Thus, we take the average between the quality of the district courts of the seller and buyer, to obtain measures $\overline{days^{first}}(\nu,\omega) = \frac{days^{first}(\nu)+days^{first}(\omega)}{2}$ and $\overline{days^{decision}}(\nu,\omega) = \frac{days^{decision}(\nu)+days^{decision}(\omega)}{2}$.

Lastly, to simplify our analysis, we create a discrete measure of pairwise court quality, where we define dummy variables $\mathbb{I}^{court,first}(\nu,\omega)$ and $\mathbb{I}^{court,decision}(\nu,\omega)$, which are equal to 1 when $\overline{days^{first}}(\nu,\omega)$ and $\overline{days^{decision}}(\nu,\omega)$ belong to the upper quartile of the distribution, respectively. Then, the pairs of worst-performing courts have a dummy variable equal to 1.

With these definitions, we turn to estimate the following specification, first for a sample of only differentiated products, and then for a sample of exclusively non-differentiated products:

$$\begin{aligned} \ln n_g(\nu,\omega,t) = \quad & \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu,\omega) + \xi_1 \left(BC(\nu,\omega) \times \mathbb{I}^{court}(\nu,\omega)\right) \\ & + \eta \Xi(\nu,\omega) + \epsilon_g(\nu,\omega), \end{aligned} \tag{5}$$

Table 5 shows the results of estimating Equation 5. Columns 1 and 7 show that cultural proximity is important overall for differentiated goods. In contrast, we do not find such an effect in Columns 4 and 7 for non-differentiated goods. This result echoes those of Table 4.

However, in Columns 2, 3, 8, and 9, we find that the impact of cultural proximity on trade is particularly pronounced in places with court congestion for differentiated products. However, this effect of court quality is absent for non-differentiated products, as Columns 5, 6, 11, and 12 show.

We interpret this as a suggestion that firms rely on cultural proximity as a source of trust in places where institutions do not work well. In other words, cultural proximity acts as an informal substitute for the trust that a well-functioning contract would provide, as argued by Nunn (2007). Differentiated goods are branded and specific to certain producing firms. In a country with market imperfections, firms can easily renege on their commitments. This is particularly exacerbated in regions with poor court quality and low contract enforcement. Unlike homogeneous goods, firms in differentiated good markets are not easily replaceable. As a result, firms buying or selling differentiated goods will only trade with firms they know and trust, and perhaps are culturally close.[15] As we show, this is particularly exacerbated in areas where the court system is delayed and backed up.

Complexity. To further investigate how differentiated and complex products are particularly reliant on caste-based trade, we analyze how the cultural proximity results vary by the number of varieties of goods bought and sold by firms (Levchenko, 2007). We first count how many 4-digit HS codes a firm buys or sells. Here, the more varieties of goods a firm buys or sells, the more complex it is. Then we estimate

$$\ln y\left(\nu, \omega\right) = \iota_{\nu} + \iota_{\omega} + \delta BC\left(\nu, \omega\right) + \eta \Xi\left(\nu, \omega\right) + \zeta_0 var + \zeta_1 \left(BC\left(\nu, \omega\right) \times var\right) + \varepsilon\left(\nu, \omega\right), \quad (6)$$

where var_{ν}^{sold} and var_{ν}^{bought} refer to the number of varieties sold and bought by the seller, while var_{ω}^{sold} and var_{ω}^{bought} refer to the number of varieties sold and bought by the buyer.

Table 6 shows that the effects of cultural proximity on trade are stronger when firms buy and sell more varieties. Our interpretation of these findings is that firms that buy or sell more varieties of goods (i.e., they are more complex firms) have to face more contracting frictions, caused by having to negotiate more contracts. Then, these firms, to minimize their load of contracting frictions, will rely more on trading with counterparts that they trust. Moreover, this explanation based on trust is compatible with the results related to differentiated goods and contract enforcement. In both cases, we posit that the intensity of trade is driven by trust between firms, to overcome market imperfections in India.

Cancellations. Our data is unique in that it records canceled transactions as well. Among the diverse reasons for which cultural proximity could affect trade, we can also study reneged contracts. In this section, we analyze whether it becomes more likely pairs of firms will cancel their transactions if they are far in cultural terms. We estimate the following specification:

$$\mathbb{I}_{g}^{cancel}\left(\nu, \omega, t\right) = \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC\left(\nu, \omega\right) + \eta \ln dist\left(\nu, \omega\right) + \epsilon_{g}\left(\nu, \omega, t\right), \quad (7)$$

where $\mathbb{I}_{g}^{cancel}\left(\nu, \omega, t\right)$ is an indicator that says if there was at least one canceled transaction going from firm ν to firm ω of good g (at the 6-digit HS classification) in month t, $\iota_{\nu g}$ is a seller-good fixed effect, ι_{gt} is a good-month fixed effect and ι_{ω} is a seller-level fixed effect. Here, we control for the month of the year to account for macro events that could have caused widespread cancellations.

[15] This relates to Rauch (1999), who mentions that search frictions (i.e. having to look for a trustworthy supplier) are more important to the trade of differentiated goods than to the trade of non-differentiated goods.

Table 5: Effect of cultural proximity on trade interacted by court quality, intensive margin

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales
Diff. good classification	Conserv.	Conserv.	Conserv.	Conserv.	Conserv.	Conserv.
Type of good	Diff.	Diff.	Diff.	Non-diff.	Non-diff.	Non-diff.
BC	0.093***	0.044	0.053	0.023	0.020	-0.005
	(0.030)	(0.036)	(0.035)	(0.055)	(0.039)	(0.035)
$BC \times \mathbb{I}^{court,first}$		0.219**			0.008	
		(0.100)			(0.133)	
$BC \times \mathbb{I}^{court,decision}$			0.169*			0.066
			(0.097)			(0.149)
Obs.	122,793	122,793	122,793	46,735	46,735	46,735
Adj. R2	0.837	0.837	0.837	0.849	0.849	0.849
FE	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales	Log Sales
Diff. good classification	Liberal	Liberal	Liberal	Liberal	Liberal	Liberal
Type of good	Diff.	Diff.	Diff.	Non-diff.	Non-diff.	Non-diff.
BC	0.104***	0.045	0.051	-0.029	-0.050	-0.060*
	(0.033)	(0.040)	(0.039)	(0.047)	(0.036)	(0.035)
$BC \times \mathbb{I}^{court,first}$		0.283**			0.066	
		(0.120)			(0.107)	
$BC \times \mathbb{I}^{court,decision}$			0.227**			0.091
			(0.113)			(0.116)
Obs.	115,459	115,459	115,459	54,000	54,000	54,000
Adj. R2	0.830	0.830	0.830	0.853	0.853	0.853
FE	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating Equation 5. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Columns 1 to 3 shows results for differentiated goods according to the conservative classification of Rauch (1999). Columns 4 to 6 show results for non-differentiated goods according to the conservative classification of Rauch (1999). Columns 7 to 9 shows results for differentiated goods according to the liberal classification of Rauch (1999). Columns 10 to 12 show results for non-differentiated goods according to the liberal classification of Rauch (1999). Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and 4-digit HS level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019). $\mathbb{I}^{court,first}$ indicates if the mean of the average number of days in between a court filing and first hearing for the origin-district court and the destination-district court is above the 75th percentile of the distribution. $\mathbb{I}^{court,decision}$ indicates if the mean of the average number of days in between a court filing and decision for the origin-district court and the destination-district court is above the 75th percentile of the distribution.

Table 6: Effect of cultural proximity on trade by number of varieties, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.111***	0.090**	0.107***	0.097**
	(0.040)	(0.040)	(0.035)	(0.039)
$BC \times var_{\nu}^{sold}$	0.089			
	(0.126)			
$BC \times var_{\nu}^{bought}$		0.121		
		(0.084)		
$BC \times var_{\omega}^{sold}$			0.112**	
			(0.051)	
$BC \times var_{\omega}^{bought}$				0.068
				(0.043)
Obs.	32,843	32,843	32,843	32,843
Adj. R2	0.410	0.410	0.410	0.410
FE	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.

	(5)	(6)	(7)	(8)
Dep. Variable	Log Transactions	Log Transactions	Log Transactions	Log Transactions
BC	0.056*	0.030	0.056*	0.042
	(0.032)	(0.032)	(0.029)	(0.032)
$BC \times var_{\nu}^{sold}$	0.095			
	(0.105)			
$BC \times var_{\nu}^{bought}$		0.141**		
		(0.067)		
$BC \times var_{\omega}^{sold}$			0.104**	
			(0.042)	
$BC \times var_{\omega}^{bought}$				0.071**
				(0.036)
Obs.	32,843	32,843	32,843	32,843
Adj. R2	0.356	0.357	0.357	0.357
FE	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.

Notes: This table shows the results of estimating Equation 6. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. $varieties_{\nu}^{sold}$ and $varieties_{\nu}^{bought}$ refer to the number of different HS codes at the 4-digit level sold and bought by the seller divided by 100, respectively. $varieties_{\omega}^{sold}$ and $varieties_{\omega}^{bought}$ refer to the number of different HS codes at the 4-digit level sold and bought by the buyer divided by 100, respectively.

Table 7 presents the results. We find that the closer firms are in cultural terms, the less likely it is that there will be a cancellation. Here we must highlight that cancellations can occur for reasons other than reneged contracts.

Table 7: Cancellations

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Cancellation indicator (0/1)					
BC	-0.006*	-0.006*	-0.006*	-0.005*	-0.000	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Obs.	252,191	252,191	248,192	256,819	258,481	252,829
Adj. R2	0.102	0.102	0.110	0.102	0.0695	0.108
FE	Seller×HS, buyer	Seller×HS, buyer, month	Seller×HS, buyer, month×HS	Seller×HS, buyer, origin×dest.	Seller×HS, buyer, month, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating Equation 7 at the transaction level. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Dependent variable is a dummy variable that is equal to 1 if there was at least 1 cancellation of a transaction involving good g for a given pair of seller and buyer in month t. Good g is defined according to 6-digit HS classification. Sales are trimmed by 4-digit HS code at 5 and 95 percent. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and 4-digit HS level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

3.1.2 Preference-based mechanisms and discrimination.

Social hierarchies. To investigate the importance of vertical hierarchies and discrimination across cultural groups, we study whether there are asymmetric effects in transactions in which one firm is placed higher than the other based on the Varna-based hierarchy. This is one way of testing for preference-based discrimination across the social hierarchy. We define the Varna or religion for which a firm has the highest probability of belonging to.[16] We do not find evidence that hierarchies (and preference-based discrimination) across social groups matter for our cultural proximity results.

We use of two different indicators: $\mathbb{I}_{\nu_H \omega_L}$ and $\mathbb{I}_{\nu_L \omega_H}$. The first one captures that the seller belongs to a higher hierarchy than the buyer. The second indicates the seller is placed below the buyer in the social hierarchy. We include these two indicators by interacting them with our measure of cultural proximity.

[16]While the Varna-based hierarchy only relates to the Hindu religion, we also place other religions in this hierarchy based on their income levels.

For the intensive margin, we estimate

$$\begin{aligned}\ln y(\nu,\omega) = \iota_\nu + \iota_\omega + \delta BC(\nu,\omega) + \eta \Xi(\nu,\omega) + \zeta_1 \left(BC(\nu,\omega) \times \mathbb{I}_{\nu_H \omega_L}\right) \\ + \zeta_2 \left(BC(\nu,\omega) \times \mathbb{I}_{\nu_L \omega_H}\right) + \varepsilon(\nu,\omega),\end{aligned} \tag{8}$$

while for the extensive margin we estimate

$$\begin{aligned}tr(\nu,\omega) = \iota_\nu + \iota_\omega + \delta BC(\nu,\omega) + \eta \Xi(\nu,\omega) + \zeta_1 \left(BC(\nu,\omega) \times \mathbb{I}_{\nu_H \omega_L}\right) \\ + \zeta_2 \left(BC(\nu,\omega) \times \mathbb{I}_{\nu_L \omega_H}\right) + \varepsilon(\nu,\omega,t).\end{aligned} \tag{9}$$

Table 8 presents the results for the intensive and extensive margins. The baseline category is that both firms belong to the same hierarchy. First, we find that the baseline coefficient is very similar to those of Table 1. Second, we find that there is no additional effect of cultural proximity when firms are placed differently in the hierarchy. We conclude that strong asymmetric effects caused by vertical discrimination across cultural groups are unlikely. The effect of cultural proximity is similar, whether or not the firms trading belong to the same or different hierarchies.

Age of firms. Preference-based discrimination is not profit maximizing, and is likely to lead to discriminating firms exiting the market. We follow Becker (1957) to analyze whether taste-based discrimination is behind our main findings. If there is taste-based discrimination, then we should see that firms that sell to culturally close firms at lower prices are willing to forego profits because of their preferences. A consequence would be that these firms are more prone to go bankrupt.

For our empirical analysis, we leverage information on the establishment date from IndiaMART (the date on which a firm was established) and registration date from the tax authority (the date on which a firm obtained its permit to trade). If there is taste-based discrimination, then we should see older firms relying less on cultural proximity. This would mean that firms that had a preference for selling to firms culturally close to them eventually went bankrupt, while the survivors were those firms that did not show these preferences. Here, we estimate

$$\ln y(\nu,\omega) = \iota_\nu + \iota_\omega + \delta BC(\nu,\omega) + \eta \Xi(\nu,\omega) + \zeta \left(BC(\nu,\omega) \times \ln\left(age(\nu)\right)\right) + \varepsilon(\nu,\omega), \tag{10}$$

where $age(\nu)$ is the age of seller ν.

Table 9 shows the results for a modified version of the intensive margin regressions according

Table 8: Effect of cultural proximity on trade by vertical hierarchies, intensive and extensive margins

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions	Trade Indicator	Trade Indicator
BC	0.099***	0.068**	0.129***	0.079***	0.0010***	0.0010***
	(0.034)	(0.028)	(0.035)	(0.029)	(0.0001)	(0.0001)
$BC \times \mathbb{I}_{\nu_H \omega_L}$	0.023	0.097	0.008	0.072	-0.0002	-0.0003
	(0.113)	(0.091)	(0.116)	(0.092)	(0.0003)	(0.0003)
$BC \times \mathbb{I}_{\nu_L \omega_H}$	0.045	-0.076	-0.027	-0.123	-0.0002	-0.0004
	(0.128)	(0.102)	(0.129)	(0.103)	(0.0002)	(0.0002)
Obs.	30,997	30,997	31,119	31,119	5,456,512	5,477,548
Adj. R2	0.418	0.360	0.412	0.357	0.614	0.0107
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation 8. Columns 5 and 6 show the results of estimating Equation 9. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer in the year 2019, and it is equal to 0 otherwise. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019). The subindex that accompanies ν denotes the hierarchical position of the seller, while the subindex that accompanies ω denotes the hierarchical position of the buyer. H denotes a higher position and L denotes a lower position. The baseline category is when both firms have the same hierarchical position.

to Equation 1. If there was taste-based discrimination, then the interaction between the measure of cultural proximity and age should have a negative coefficient. However, we find weak evidence for taste-based discrimination, such that we cannot conclude this is the reason behind our results.

3.1.3 Cultural-group cartels

We explore whether our empirical patterns can be rationalized by a narrative of certain cultural-group cartels controlling a certain sector or market location, and excludes outsiders. This would imply different normative conclusions, as stabilizing cartels would lead to an efficiency loss. While not widely highlighted in the literature, certain case studies exist that may tangentially support these narratives. Such as the example of the Gounders community's specialized knitted garment production in Tirupur (Munshi, 2019).

Table 9: Effect of cultural proximity after controlling for establishment age of sellers, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
A) IndiaMART				
BC	0.734**	0.489*	0.800**	0.479
	(0.355)	(0.296)	(0.371)	(0.311)
$BC \times \ln(age(\nu))$	-0.199*	-0.124	-0.207*	-0.122
	(0.111)	(0.090)	(0.112)	(0.091)
Obs.	6,334	6,334	5,859	5,859
Adj. R2	0.428	0.303	0.387	0.237
B) Tax Authority				
BC	0.164	0.217**	0.150	0.233**
	(0.115)	(0.094)	(0.116)	(0.097)
$BC \times \ln(age(\nu))$	-0.032	-0.076*	-0.016	-0.082**
	(0.050)	(0.041)	(0.050)	(0.041)
Obs.	18,268	18,268	18,810	18,810
Adj. R2	0.406	0.333	0.403	0.332
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation 10. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019). In Panel A, the age of the seller comes from data reported in IndiaMART. In Panel B, the age of the seller comes from data reported by the tax authority.

Here, we first construct the Herfindahl-Hirschman index (HHI) of cultural-group concentration, and test its importance in moderating the impact of cultural proximity. We measure the HHI across castes at the product level HHI^{good}, and the location level $HHI^{district}$. Then, we assign each seller ν to their corresponding caste's HHI. The idea is to test if products or places with more market concentration (where cartels may survive and be important) have a greater positive impact of cultural proximity.

Furthermore, in order to simplify the interpretation of our results, we define the dummy variable $\mathbb{I}^{HHI}(\nu)$, which is equal to 1 when $HHI(\nu)$ belongs to the upper half of the distribution. That is, the dummy variable will be equal to 1 for those firms that belong to the castes with a higher HHI.

Then, we estimate the following regression:

$$\begin{aligned}\ln n_g(\nu,\omega,t) = \quad & \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu,\omega) + \xi\left(BC(\nu,\omega)\times \mathbb{I}^{HHI}(\nu)\right) \\ & + \eta \Xi(\nu,\omega) + \epsilon_g(\nu,\omega)\,.\end{aligned} \tag{11}$$

Table 10 shows that market concentration has little effect on the impact of cultural proximity on sales. While we find little evidence of cartels driving our effects, we are unable to rule out that such forces may not be at play for certain products and locations.

Table 10: Effect of cultural proximity on trade interacted with market concentration, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.069**	0.083***	0.062**	0.078***
	(0.032)	(0.030)	(0.030)	(0.027)
$BC \times \mathbb{I}^{HHI,good}$	0.006		0.010	
	(0.046)		(0.049)	
$BC \times \mathbb{I}^{HHI,district}$		-0.043		-0.049
		(0.059)		(0.051)
Obs.	226,039	225,775	229,719	229,719
Adj. R2	0.853	0.853	0.854	0.854
FE	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating Equation 11. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and 4-digit HS level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019). $\mathbb{I}^{HHI,good}$ indicates the caste of the seller is in the upper half of the good-based HHI distribution. $\mathbb{I}^{HHI,district}$ indicates the caste of the seller is in the upper half of the district-based HHI distribution.

3.2 Discussion of stylized facts

The stylized facts show that higher cultural proximity between a pair of firms favors trade in both the intensive and extensive margins, as well as lowers the price of the goods they trade. We discuss the possible mechanisms that may give rise to these findings.

Contracting frictions. In Section 3.1.1, we argue that contracting frictions could be the reason that drives the cultural proximity results. India is a country that suffers from a severe lack of contract enforcement. *A priori*, a buyer may not know if the seller will deliver the goods under the agreed conditions (delivery, quality, etc.). Likewise, *a priori*, the seller may not know if the buyer will pay under the agreed conditions. This means buyers and sellers incur contracting frictions to find suitable trading partners (Boehm and Oberfield, 2020). Quantity-wise and matching-wise, this lowers trade as firms must pay a matching cost. Price-wise, this increases prices as the matching cost is passed down by the sellers to the aforementioned prices.

In this case, cultural proximity can work as a proxy for information and trust: culturally close firms may know and trust each other, and informally enforce contracts with social and reputational pressures. The higher the cultural proximity, the lower the contracting frictions. As a result, there would be more trade and lower prices, consistent with our previous findings. In Section 4.1, we present a simple theoretical framework in which cultural proximity affects contracting frictions and affects trade and prices. Our model suggests that if contracting frictions drive initial trade barriers, cultural proximity may reduce such frictions.

Preference-based mechanisms and discrimination. We argue that the results are unlikely to emerge from buyers having an inherent preference for buying from sellers culturally close to them. This preference would be a demand shifter that is active for those sellers that are close in cultural terms. While this would certainly increase the quantity traded, it would increase the price of traded goods, a result inconsistent with our previous findings.

The stylized facts can arise from having sellers who show a preference for selling to culturally close buyers. This would imply increased supply for those buyers who are culturally close to the seller. However, we do not find conclusive evidence of this channel.

Discrimination from high-caste cultural groups against low-caste cultural groups may again reduce trade. Yet, we find there is no additional effect of cultural proximity when firms are placed differently in the hierarchy. As such, we detect no asymmetric effects caused by vertical discrimination across cultural groups.

Cultural-group cartels. While we cannot, conclusively rule out the role of cultural-group cartels, in Table 10 we find little evidence in support of it.

We begin with the empirical assumption that cartels are more likely to have market power on a particular set of products or locations. It is challenging for a cartel to exist in a broad variety of industries and locations all over the country; as they may need substantial market power in many industries and regions. Yet, in Table 6 we show that, if anything, our results are stronger when more products are transacted – perhaps as more products imply more contracting frictions.

Furthermore, if cartels specialize in producing certain goods, in Appendix Table A7 we find evidence that the effects of cultural proximity are not particularly stronger when firms sell the goods in which their own castes specialize in. Relatedly, even if the cartel is not focused on certain products, it may have a presence in a broader industry. But in controlling for seller-industry-by-buyer-industry fixed effects, our results are similarly robust, as Appendix Table A8 shows.

In a similar vein, it is unlikely that these cartels emerge from regional concentration and market power, as our regressions include origin-by-destination fixed effects, which implies that we are looking within location pairs.

Lastly, our results are also inconsistent with the insight that cartels often consist of entrenched old firms. Table 9 shows that, for older firms, cultural proximity matters less, perhaps as they have smoothed out contracting frictions over time.

3.3 Robustness

We examine alternative specifications and heterogeneity in responses that sheds light on various other channels in Appendix B.

Correction for selection bias. Helpman et al. (2008) shows that the standard gravity equation estimations are biased as they do not account for selection issues. We follow their suggested correction in Appendix Tables A2 and A3. As the correction mentions, we need an excluded instrument that affects only the extensive margin (i.e., the matching cost) and not the intensive margin (i.e., the trade cost). We consider the participation of both seller and buyer in the IndiaMART online B2B platform, under the idea that online platforms should reduce their matching costs. The results show that the coefficients are downward-biased if we do not account for the selection issues. Therefore, our main results in the paper provide a lower bound on the effect that cultural proximity has on the intensive margin of trade.

Alternative cultural proximity measure. As an alternative to the Bhattacharyya coefficient, we perform estimation exercises using a symmetric version of the Kullback and Leibler

(1951) divergence. Appendix Tables A4 and A5 show our baseline findings are robust to this alternative cultural proximity measure.

Language. We test whether the results we find are driven by linguistic similarity. To do so, we follow the two linguistic distance measures from Kone et al. (2018). Appendix Table A6 shows that, after controlling for linguistic similarity between firms, cultural proximity remains statistically significant to explain inter-firm trade. This result suggests that cultural proximity encompasses cultural factors beyond language barriers that explain trade at the micro level.

Goods specialization. Cultural groups in India are, in many cases, defined by the production of specific goods (Munshi, 2019).[17] Therefore, we analyze if the reason behind the cultural proximity results is cultural groups specializing in the production of certain goods and, given this, forming special bonds with their specific set of buyers. In Appendix Table A7, we do not find evidence of good specialization driving the results. This means that cultural proximity matters for all types of goods: for those in which a cultural group specializes, and for those in which a cultural group does not specialize too.

Industry pair linkages. In the production matrix of an economy, some sectors are more likely to trade with others because of the nature of their activities. The same cultural group may happen to participate in the same industry. In Appendix Table A8, we present the results for the intensive margin after adding an industry of seller $\times$ industry of buyer fixed effect. We find that the result of there being more trade between culturally close firms prevails.

We end this section by mentioning that last names are not interesting by themselves, nor do they necessarily directly affect behavior of the firms. However, while firms are not human entities, they are composed of people and owners who engage with the people and owners of other firms. As such, names are relevant as a signal of which cultural group an owner belongs to, being especially useful when it comes to people forming relationships with others. As firms are composed of people, names end up affecting firm outcomes.

This is especially true in our case, where firms in our dataset are mostly small (see Appendix Figure A1), so they rely on relationship-based contracting. Because the names in India are a strong signal of caste, religion, and community group, knowing them allows owners to form relationships and enforce contracts informally outside the court system.

[17]We can also understand this as certain cultural groups specializing in certain occupations.

4 Quantitative Importance of Cultural Proximity

In this section, we introduce a theoretical model of firm networks with cultural proximity and perform a counterfactual analysis to quantify the importance of cultural proximity for trade and welfare. First, we describe the model, which is based on that of Bernard et al. (2022). Then, we describe how we estimate and calibrate the model. Finally, we perform three counterfactual scenarios. In particular, we show the importance of cultural proximity by implementing (i) social inclusion/mixing policies, (ii) social isolation policies, and (iii) reduction of contracting frictions.

4.1 Model

4.1.1 Summary

We build a quantitative inter-firm trade model and cultural differences between firm owners. The economy is comprised of a representative household and a continuum of firms. The representative household has preferences over all goods in the economy. Goods are differentiated and produced by the firms. Each good can be consumed by the household as a final good, or purchased as an intermediate input by other firms. Trade between the household and firms is riskless, but trade between firms is risky, such that the equilibrium price between firms is subject to a risk premium. Finally, we consider an endogenous firm network, where sellers optimally match with buyers, ensuring that the sellers' profits from matching are greater than the matching costs. Following our stylized facts, both premium and matching costs are a function of the cultural proximity between firms. For details on the derivations of the model see Appendix C.

4.1.2 Technology

A continuum of firms that operate under monopolistic competition and produce differentiated goods indexed by ω. Each firm produces a unique good, so ω denotes both a firm and a good. Each firm produces its good using labor from the representative household and intermediate inputs from others firms in the economy.

Firms operate in three steps. First, sellers endogenously choose buyers (i.e. matching).[18] In particular, a seller ν sells to buyer ω whenever the profits of doing so are larger than

[18] Following Bernard et al. (2022), we assume that sellers bear all the fixed costs of matching with buyers. The primary benefit of this assumption is tractability, as sellers' decisions to match with each buyer are separable. Similarly, assuming that buyers bear all the fixed costs of matching involves buyers solving complicated combinatorial sourcing problems as in Antras et al. (2017). Even though relevant in other contexts, focusing on interdependent sourcing problems is outside of the scope of our paper.

the fixed costs of matching $\epsilon F(\nu,\omega)$. Firms meet and the seller pays the fixed cost $F(\nu,\omega)$ upon matching. This term includes costs that sellers must incur to identify reliable clients, similar to De Sousa et al. (2023). In our setup, culture encodes information that is useful for sellers to decide which buyers to sell to (Allen et al., 2019; Balmaceda and Escobar, 2017). Then, guided by our stylized facts, cultural proximity $BC(\nu,\omega)$ between seller ν and buyer ω determines the pairwise fixed matching costs $F(\nu,\omega)$.[19]

Second, upon matching, seller ν and buyer ω set up the terms of the contract. There is a risk that the buyer withholds payment, so the seller charges a risk premium $d(\nu,\omega) \geq 1$ to the buyer. Guided by our stylized facts, we posit that cultural proximity $BC(\nu,\omega)$ between seller ν and buyer ω determines $d(\nu,\omega)$. Third, sellers and buyers trade.

Step 1: Matching. In the first step, we endogenize sellers selecting which buyers to trade with by laying out the maximization problem of firms, and how cultural proximity influences it. A seller ν matches with buyer ω whenever the seller's profits of doing so $\pi(\nu,\omega)$ are larger than the fixed costs of matching $\epsilon F(\nu,\omega)$, where $F(\nu,\omega)$ are pairwise fixed costs of matching, and ϵ are i.i.d. log normal errors with mean $\mu_{\ln(\epsilon)}$ and standard deviation $\sigma_{\ln(\epsilon)}$. It can be shown that profits are proportional to the value of intermediate sales $n(\nu,\omega)$ from seller ν to buyer ω, such that

$$\pi(\nu,\omega) = \frac{n(\nu,\omega)}{\sigma}. \tag{12}$$

Sellers are looking for buyers, and vice-versa. Firms meet to infer the fixed costs the seller would incur if they were to trade with each buyer. Each firm owner has a name, and sellers and buyers exchange names upon meeting. Then, each name has a probability mapping to different cultural groups. We assume that this mapping is public information.[20]

For each cultural group j, we consider the matching function $M^j(\nu,\omega) = (\rho_\nu(j))^\varphi (\rho_\omega(j))^{1-\varphi}$, where $\rho_\nu(j)$ is the probability that seller ν belongs to cultural group j, $\rho_\omega(j)$ is the probability that buyer ω belongs to cultural group j, φ is the weight of ν to determine their proximity. Asssuming $\varphi = \frac{1}{2}$, the expected proximity between ν and ω is proportional to the Bhattacharyya coefficient we used to measure cultural proximity between firms: $\overline{M}(\nu,\omega) = \frac{1}{\mathcal{X}}\sum_{j=1}^{\mathcal{X}} M^j(\nu,\omega) = \frac{1}{\mathcal{X}}\sum_{j=1}^{\mathcal{X}} \sqrt{\rho_\nu(j)\,\rho_\omega(j)} = \frac{1}{\mathcal{X}} BC(\nu,\omega)$.

Finally, following our stylized fact on how the Bhattacharyya coefficients influence inter-firm

[19] There could be other microfoundations motivating how culture influences matching costs (e.g., risk-sharing as in Ambrus et al. 2014; Bloch et al. 2008). These would generate similar frictions at the extensive and intensive margin.

[20] For example, all firm owners know that the surname *Shah* is associated with either of the three groups: Jain, Muslim (Faqir), or Hindu (Vaishnav baniya), with probabilities reflecting the empirical distribution.

trade at the extensive margin, we consider the fixed costs of matching to be a function of cultural proximity such that $F(\nu,\omega) = f\left(\overline{M}(\nu,\omega)\right)$, where $f(\cdot)$ is continuous function. Our modeling decision allows for the fact that a firm belonging to a cultural group encodes information that other firms use to determine the cost of matching. For computational and analytical convenience, we consider an exponential function for matching costs:

$$F(\nu,\omega) = \kappa + \exp\left(\gamma BC(\nu,\omega)\right), \tag{13}$$

where γ is the semi-elasticity of the matching cost to cultural proximity, and κ is a scaling constant. γ captures the sensitivity of matching costs $F(\nu,\omega)$ to cultural proximity $BC(\nu,\omega)$, including institutional features such as court quality. That is, the lower the quality of courts is, the larger the parameter γ is, and the more firms rely on cultural proximity to match. This goes in line with current research as in Boehm et al. (2024).[21]

We now describe how firms match. Before the formation of the network, firms are characterized by $\lambda = (z, \boldsymbol{\rho})$, where z is productivity, and $\boldsymbol{\rho}$ is the vector of probabilities of firm λ belonging to each cultural group. After firms meet each other and infer their cultural proximity, firms now only differ in their productivity z. So the share of seller-buyer pairs $\left(z, z'\right)$ is

$$l\left(z,z'\right) = \int \mathbb{I}\left[\ln\left(\pi\left(z,z'\right)\right) - \ln\left(F\left(z,z'\right)\right) - \ln\left(\epsilon\left(z,z'\right)\right) > 0\right] dH\left(\epsilon\left(z,z'\right)\right), \tag{14}$$

where $l\left(z,z'\right)$ is called the *link function*. From Equations 13 and 14, we see that the higher the cultural proximity, the lower the matching cost and the larger the probability of matching. This relates to Stylized Fact 2.

Step 2: Terms of the contract. After matching, seller ν and buyer ω set up the terms of the contract. We consider the possibility that the buyer may withhold payment after the seller ships the goods. Then, the seller charges a risk premium to the buyer. This premium gets passed to the unit price $p(\nu,\omega)$ that results from the seller maximizing profits given demand from other firms:

$$p(\nu,\omega) = \mu c(\nu)\, d(\nu,\omega), \tag{15}$$

where $c(\nu)$ is the marginal cost and $\mu \equiv \frac{\sigma}{\sigma-1}$ is a markup. We now turn to explain the premium $d(\nu,\omega)$. Consider that the settlement of the transaction happens within a continuous

[21] In the context of India, Boehm et al. (2024) show that higher contracting frictions (measured by lower court quality) leads to more sticky relationships between firms (i.e., higher fixed costs of matching). They also show that this leads to firms relying more on relational agreements (e.g., cultural proximity).

unit of time. Seller ν considers the time T that buyer ω withholds payment for the first time is a random variable, where T follows an exponential distribution with intensity rate $\delta(\nu,\omega)$. Then, the probability seller ν waits for more than t units of time until buyer ω withholds payment for the first time is $P(T>t)=\exp(-\delta(\nu,\omega)t)$. Seller ν cares that buyer ω never withholds payment. Then, the probability $\overline{p}$ that seller ν waits more than a unit of time until buyer ω withholds payment is $\overline{p}=P(T>1)=\exp(-\delta(\nu,\omega))$. Then, the expected number of times buyer ω withholds payment is $\frac{1}{\overline{p}}=\exp(\delta(\nu,\omega))$. We posit that the premium $d(\nu,\omega)$ is proportional to the expected number of times buyer ω withholds payment to seller ν. The intuition is that seller ν will charge a higher premium if there is a higher hold-up risk from buyer ω. For simplicity, we consider $d(\nu,\omega)=\frac{1}{\overline{p}}=\exp(\delta(\nu,\omega))$.

Finally, the intensity rate $\delta(\nu,\omega)$ includes a set of bilateral factors that influence trade the risk premium from seller ν to buyer ω. This premium can include geographic or institutional factors. Guided by our stylized fact on how cultural proximity shapes inter-firm trade, we consider the Bhattacharyya coefficient $BC(\nu,\omega)$ to be part of $\delta(\nu,\omega)$, such that

$$d(\nu,\omega)=\exp(\beta BC(\nu,\omega)+\epsilon(\nu,\omega)), \tag{16}$$

where β is the semi-elasticity of the risk premium to cultural proximity, and $\epsilon(\nu,\omega)$ are unobservables.[22] β captures sensitivity of the risk premium $d(\nu,\omega)$ to cultural proximity $BC(\nu,\omega)$ including institutional features such as court quality.[23]

Step 3: Trade. Finally, firms engage in trade. Each firm has a technology $y(\omega)=\kappa_\alpha z(\omega)\,l(\omega)^\alpha\,m(\omega)^{1-\alpha}$, where $y(\omega)$ is output, $\kappa_\alpha\equiv\frac{1}{\alpha^\alpha(1-\alpha)^{1-\alpha}}$ is a normalization constant, $z(\omega)$ is firm-level productivity, $l(\omega)$ is labor, and $m(\omega)$ are intermediate inputs from other firms. In turn, the intermediate inputs are defined as a CES composite $m(\omega)=\left(\int_{\nu\in\Omega(\omega)}m(\nu,\omega)^{\frac{\sigma-1}{\sigma}}d\nu\right)^{\frac{\sigma}{\sigma-1}}$, where $m(\nu,\omega)$ is quantity of inputs from seller ν to buyer ω, $\sigma>1$ is the elasticity of substitution across intermediates, and $\Omega(\omega)$ is the endogenous set of suppliers of buyer ω.

Through cost minimization, firms' unit cost function is $c(\omega)=\frac{P(\omega)^{1-\alpha}}{z(\omega)}$, where $P(\omega)\equiv\left(\int_{\nu\in\Omega(\omega)}p(\nu,\omega)^{1-\sigma}d\nu\right)^{\frac{1}{1-\sigma}}$ is a CES price index across prices of intermediates, and labor is the numeraire good, so $w=1$. Profit maximization subject to demand generates constant

[22] This is similar to earlier work on how culture can be an informal institutional channel to solve hold-up problems. This would be consistent with other microfoundations on how culture influences trade costs due to reputation (Banerjee and Duflo, 2000; Chen and Wu, 2021) or loyalty (Board, 2011).

[23] Based on the result of Table 5, we know model parameter β increases in magnitude when $\mathbb{I}^{court}=1$ and firms face bad courts. As such, β relates to how firms respond to court quality: the lower the quality of the courts, the larger the magnitude of β, and the more important cultural proximity is for trade.

markup pricing such that the unit price is $p(\nu,\omega) = \mu c(\nu)\, d(\nu,\omega)$, and the demand for intermediates is $n(\nu,\omega) = p(\nu,\omega)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega)$, where $N(\omega) = \int_{\nu\in\Omega(\omega)} n(\nu,\omega)\, d\nu$ is the total intermediate purchases by buyer ω and $n(\nu,\omega) \equiv p(\nu,\omega)\, m(\nu,\omega)$ is the value of purchases from seller ν to buyer ω. From the demand of intermediates and firm pricing, we can obtain the gravity equation as

$$\log(n(\nu,\omega)) = \iota + \iota_\nu + \iota_\omega + (1-\sigma)\log(d(\nu,\omega)), \tag{17}$$

where $\iota \equiv \log(\mu^{1-\sigma})$, $\iota_\nu \equiv \log\left(c(\nu)^{1-\sigma}\right)$, and $\iota_\omega \equiv \log\left(P(\omega)^{\sigma-1} N(\omega)\right)$. Here, the premium $d(\nu,\omega)$ enters the gravity equation as a trade cost. This gravity equation relates directly to Equation 1 that we estimate.

4.1.3 Preferences

A representative household demands goods from all firms, inelastically supplies a unit of labor to firms, and owns the firms. The household maximizes utility $\left(\int_{\omega\in\Omega} y(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right)^{\frac{\sigma}{\sigma-1}}$ subject to its budget constraint $\int_{\omega\in\Omega} q(\omega)\, y(\omega)\, d\omega \leq Y$, where $y(\omega)$ is the household demand for good ω, $q(\omega)$ is the price the household pays for good ω, and Y is total income. Since firm ω sells its good to the household without risk, the price is $q(\omega) = \mu c(\omega)$, where μ is the markup and $c(\omega)$ is firm ω's marginal cost. Finally, the household income is $Y = wL + \Pi$, where wL is labor income, and Π are total profits.[24]

4.1.4 Solving the model

Based on the price index of all of the goods acquired by firm $z^{'}$, we get

$$P\left(z^{'}\right)^{1-\sigma} = \mu^{1-\sigma} \int P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} d\left(z,z^{'}\right)^{1-\sigma} l\left(z,z^{'}\right) dG(z), \tag{18}$$

where $l\left(z,z^{'}\right)$ is the share of sellers of productivity z that sell to buyers with productivity $z^{'}$, also called the *link function*. Now, total sales of firm z is the sum of sales to households plus intermediates, so

$$S(z) = \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1}\right] \times \left[\frac{Y}{Q^{1-\sigma}} + \left(\frac{1-\alpha}{\mu}\right)\left(\int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} S(z')\right] l(z,z')\, dG(z')\right)\right]. \tag{19}$$

[24]Although we do not have explicit trade costs or matching costs between firms and the household, our model implicitly accounts for them. Our main model assumes that households cannot directly purchase from a factory but must go through a retailer, and that retailer has matching frictions with producers.

We solve the model using a fixed point algorithm (Bernard et al., 2022). First, we make a guess of the link function $l\left(z, z'\right)$. Second, we find equilibrium prices $P(z)$ by solving Equation 18. Third, given prices $P(z)$, we find equilibrium total sales $S(z)$ by solving Equation 19. Fourth, we find trade flows $n(\nu, \omega)$ from Equation 17. Fifth, we find profits $\pi(\nu, \omega)$ from Equation 12. Sixth, we find an updated value for the link function $l\left(z, z'\right)$ from Equation 14. We iterate until convergence.

4.2 Estimation and calibration

We briefly explain how we estimate the key parameters of the model on cultural endowments, (intensive) trade costs, and seller matching costs. We also describe how we calibrate the remaining parameters of the model. For more details on the estimation, see Appendix D.

Cultural endowments $\boldsymbol{\rho}$. For the cultural endowments, we assume each firm ν has a probability vector $\boldsymbol{\rho_\nu} = [\rho_\nu(1), \ldots, \rho_\nu(452)]$ of belonging to each of the 452 cultural groups we observe in the data. We assume the elements of $\boldsymbol{\rho_\nu}$ are randomly drawn from a Dirichlet distribution, such that $\rho_\nu(1), \ldots, \rho_\nu(452) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_{452})$, where $\alpha_1, \ldots, \alpha_{452} > 0$ are concentration parameters.[25] Then, given the probability density of the Dirichlet distribution, we can estimate the vector of parameters cultural endowments $\boldsymbol{\rho_\nu}$ via maximum likelihood estimation.

Trade cost d. From Equation 16 we need an estimate for β. Our setup produces a clear empirical counterpart that we already estimated in the reduced-form section, conditional on high-dimensional fixed effects. So we obtain an estimate for this parameter by linking the theoretical gravity Equation 17 to the empirical gravity equation results (Column 1 from Table 1). Thus, we obtain $\beta = -0.03$.[26]

Matching cost F. From Equation 13, we need an estimate for γ. We do this in two steps. First, using the extensive margin sample, we estimate a gravity equation of sales $\ln\left[n\left(z, z'\right)\right]$ against cultural proximity $BC(z, z')$, seller and buyer fixed effects (ι_z and $\iota_{z'}$, respectively), and origin-destination fixed effects (ι_{od}), where we apply the inverse hyperbolic sine transformation to the dependent variable to not lose the cases in which there is zero trade. With this, we recover the predicted sales $\ln\widehat{[n(z, z')]}$. This variable predicts what

[25] For a given x, the higher this parameter, the more disperse the realizations of $\rho_\nu(x)$ are across firms ν.

[26] Even though the wedge also appears in the price Equation 15 of the model, we do not estimate this equation to identify β. The reason is that the price equation is not an equilibrium equation, while the gravity equation is. Also, for our simulations we add a constant to the trade cost, such that the minimum trade cost is equal to 1.

would be the sales for a pair of seller and buyer even in the case they did not actually trade in the data. Second, we combine and rearrange Equations 13 and 14, such that we can estimate γ via probit regression framework by replacing $\ln\left[n\left(z,z'\right)\right]$ with its predicted value $\ln\widehat{[n\left(z,z'\right)]}$. We find that $\gamma=-0.11$, and statistically different from 0.[27]

Calibrated parameters and SMM. We calibrate the labor cost share $\alpha=0.52$, the value reported for India for 2019 from the Penn World Tables (Feenstra et al., 2015). This value also considers the informal sector, which plays a large role in India. For the markup we use $\mu=1.34$, which is the median markup across all Indian sectors reported by De Loecker et al. (2016). This markup implies an elasticity of substitution across suppliers $\sigma=3.94$. We normalize the total number of workers $L=1$, take the nominal wage as the numeraire so $w=1$, and set the total number of firms $\mathcal{N}=1000$.

For the log-productivity distribution, we assume a mean $\mu_{\ln(z)}=0$. The remaining parameters are (i) the standard deviation of the log-productivity distribution $\sigma_{\ln(z)}$ and (ii) the mean $\mu_{\ln(\epsilon)}$, (iii) the standard deviation $\sigma_{\ln(\epsilon)}$ of the link function noise distribution and (iv) the scaling constant for the pairwise matching cost κ. We estimate these four parameters so as to match targeted moments from the data, using a simulated method of moments (SMM).

In Appendix D.2, we provide details on the SMM estimation, as well as on the targeted and untargeted moments we use. In Appendix Table A10, we present the estimated parameters. Then, in Appendix Table A11, we show that, when it comes to the targeted moments, the model can very closely replicate the empirical ones. For the untargeted moments, the model achieves a reasonably close fit to the data.

4.3 Counterfactual analysis

We now present the results of various counterfactual exercises. First, we evaluate the effects of social mixing/inclusion and isolation policies, whereby we change the cultural proximity between firms (in our model terms, changing $BC\left(z,z'\right)$). Second, we study the effects of a policy that reduces contracting frictions, such that firms rely less on cultural proximity when trading (in terms of our model, shrinking parameters β and γ).

To evaluate each scenario, we measure what occurs to various model-based statistics. Welfare is measured by real income, $\mathcal{W}=\frac{wL+\Pi}{P}$, where Π are the aggregate profits. To quantify the impact on aggregate productivity, we consider a sales-weighted average productivity measure

[27]We present the results of the estimation in Appendix Table A9. Also, for our simulations we add a constant to the matching cost, such that the minimum matching cost is equal to $1+\kappa$.

such that $\mathcal{Z} = \left(\sum_{\nu=1}^{N} \phi_\nu z_\nu^{\sigma-1}\right)^{\frac{1}{\sigma-1}}$, where ϕ_ν represents the proportion of the sales of firm ν over the total sales of the economy.

To analyze the impact on total economic activity, we measure the median change in total sales S_ν across all firms indexed by ν. Additionally, we consider the median change across all firms in the normalized intermediate sales $\widetilde{N}(\nu)/\mathcal{N}_b(\nu)$, where $\widetilde{N}(\nu)$ are the total intermediate sales of seller ν, while $\mathcal{N}_b(\nu)$ is the total number of buyers of that firm. We also show the median change across all firms in the normalized intermediate purchases $N(\omega)/\mathcal{N}_s(\omega)$, where $N(\omega)$ are the total purchases of buyer ω, and $\mathcal{N}_s(\omega)$ is its number of sellers.

To study how matching between firms is affected, we present median change across all firms in the normalized number of buyers, $\mathcal{N}_b(\nu)/\mathcal{N}$ (i.e. the outdegree), and the median change across all firms in the normalized number of sellers, $\mathcal{N}_s(\nu)/\mathcal{N}$ (i.e. the indegree). Finally, for the prices, we compare the changes in the aggregate price index $Q \equiv \left(\int_{\omega\in\Omega} q(\omega)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$.

Social inclusion and social mixing policies. We analyze the effects of social inclusion/mixing policies. There is an important literature on the role of social cohesion for economic development (Alesina and Giuliano, 2015; Alesina and Reich, 2015; Bazzi et al., 2019; Depetris-Chauvin et al., 2020; Gradstein and Justman, 2019; Ritzen et al., 2000). We tie our counterfactuals to the importance of implementing affirmative action policies with the intention of increasing cultural proximity (Alan et al., 2021; Alesina et al., 2021), particularly for India (Khanna, 2020; Munshi, 2019). For instance, affirmative action programs may help incentivize students from different cultural groups to attend the same educational institutions. If these students then go on to become owners of the firms in the future, such policies may increase cultural proximity between firms, despite the fact the owners originally belonged to different cultural groups. Similarly, affirmative action in public sector jobs may also increase connections across caste lines, as individuals from different castes now work together.

To analyze the maximum potential of this policy within our theoretical framework, we propose case Counterfactual 1 (CF1), in which all the firms belong to the same cultural group. That is, we go from the baseline to $BC\left(z, z'\right) = 1$ for all z, z', which makes the firms become the closest possible in cultural terms. In this scenario, there are no contracting frictions, as firms know and/or trust each other, and so they pay the minimum trade and matching costs.

Table 11 shows how the various model statistics change in each counterfactual relative to the baseline. In case CF1, we have that firms become the closest in cultural terms, so trade costs and matching costs are minimized to their lowest possible level. Aligned with our empirical findings, lower trade costs result in a median change of 0.40 percent in total sales, while the median changes in intermediate sales and purchases are 0.95 percent and 0.39 percent, respectively. With the lower matching costs the median change in the normalized number of buyers is 0.54 percent, while the median change in the normalized number of sellers is 0.50. Additionally, due to lower trade and matching costs, the aggregate price index decreases by 1.87 percent, and welfare increases by 1.90 percent.

Besides welfare, another aggregate measure we analyze is average productivity, which falls by 0.23 percent. Yet, average productivity masks substantial compositional changes, as these results depend on whether the less productive firms are selling more or less relative to the baseline case. We show in Table 12 that, in case CF1, when trade and matching costs decrease, the less productive firms sell more, which increases their weight in the aggregate and lowers average productivity.

Table 11: Effect of cultural proximity on aggregate outcomes (counterfactual scenarios)

	CF1: Social inclusion/mixing	CF2: Social isolation	CF3: Reducing contracting frictions
Welfare	1.90	-0.93	0.95
Ave. productivity	-0.23	0.13	-0.12
Median change of total sales	0.40	-0.21	0.20
Median change of int. sales	0.95	-0.51	0.48
Median change of int. purchases	0.39	-0.21	0.20
Median change of number of buyers	0.54	-0.29	0.27
Median change of number of sellers	0.50	-0.27	0.25
Agg. price index	-1.87	0.94	-0.94

Notes: We present the percentage gains or losses with respect to the baseline scenario. Changes shown represent average after estimating and simulating the model 100 times. CF1 is a case where all the firms belong to the same cultural group. This is, we go from the baseline to $BC\left(z, z^{'}\right) = 1$ for all $z, z^{'}$, which makes the firms to become the closest possible in cultural terms. In this scenario, there are no contracting frictions, as firms know and/or trust each other, and so they pay the minimum trade and matching costs. CF2 is a case where each firm belongs to its own cultural group. Thus, we have a case where $BC\left(z, z^{'}\right) = 0$ for all $z, z^{'}$ and $z \neq z^{'}$, which makes the firms the furthest possible in cultural terms. Under this scenario, firms incur the maximum contracting frictions, for which they pay the maximum trade cost and the maximum matching cost. CF3 is a scenario where trade and matching costs become less sensitive to cultural proximity. In this case parameters β and γ shrink by 50 percent.

Social isolation policies. Since the rise of democracy, efforts have been put in place by the Indian government to end the influence of the caste system in the modern economy (Iyer et al., 2013; Munshi, 2019). What would have happened if sociopolitical forces perpetuated the social stratification of the caste system? To analyze the maximum impact of social isolation policies, we propose case Counterfactual 2 (CF2), where we examine an extreme case in which each firm belongs to its own cultural group. Thus, we have a case where $BC\left(z, z^{'}\right) = 0$ for all $z, z^{'}$ and $z \neq z^{'}$, which makes the firms the furthest possible in cultural terms. Under this scenario, firms incur the maximum contracting frictions, for which they pay the maximum trade cost and the maximum matching cost.

Table 12: Change in sales by productivity quartiles

	CF1: Social inclusion/mixing	CF2: Social isolation	CF3: Reducing contracting frictions
1st quartile (most productive)	-0.02	0.01	-0.01
2nd quartile	0.43	-0.23	0.22
3rd quartile	0.38	-0.20	0.19
4th quartile (least productive)	0.34	-0.18	0.17

Notes: We aggregate the sales of all firms that belong to a productivity quartile and calculate their percentage variation to the baseline. Changes shown represent average after estimating and simulating the model 100 times. CF1 is a case where all the firms belong to the same cultural group. This is, we go from the baseline to $BC\left(z, z'\right) = 1$ for all z, z', which makes the firms to become the closest possible in cultural terms. In this scenario, there are no contracting frictions, as firms know and/or trust each other, and so they pay the minimum trade and matching costs. CF2 is a case where each firm belongs to its own cultural group. Thus, we have a case where $BC\left(z, z'\right) = 0$ for all z, z' and $z \neq z'$, which makes the firms the furthest possible in cultural terms. Under this scenario, firms incur the maximum contracting frictions, for which they pay the maximum trade cost and the maximum matching cost. CF3 is a scenario where trade and matching costs become less sensitive to cultural proximity. In this case parameters β and γ shrink by 50 percent.

When all firms are the furthest in cultural terms, trade costs and matching costs are the highest. Table 11 shows that in case CF2, the median change in total sales is -0.21 percent, while the aggregate price index goes up by 0.94 percent. The median change in intermediate sales is -0.51 percent, and the median change in intermediate purchases turns out to be -0.21 percent. There are also fewer matches, as the median change in the normalized number of buyers and sellers is -0.29 percent and -0.27 percent, respectively. As a result, the change in welfare is -0.93 percent. Average productivity increases by 0.13 percent, relative to the baseline. Table 12 shows that with social isolation, the most productive firms remain almost unchanged in terms of sales, while the least productive firms lose. This shrinks the weight of the least productive firms in the aggregate and, thus, drives average productivity up.

Reducing contracting frictions. Now we turn to study the effect of reducing contracting frictions. A policy that improves the quality of courts would reduce the contracting frictions firms face. In terms of our framework, this means that the trade cost and the matching cost become less sensitive to our measure of cultural proximity. Thus, in the Counterfactual 3 (CF3), we analyze a case where parameters β and γ shrink by 50 percent. This captures how reducing contracting frictions affects aggregate outcomes via the channel of trade becoming less reliant on cultural proximity.[28]

[28] Reducing contracting frictions may affect aggregate outcomes through other channels as well, such as more investments in differentiated products, and more trade across longer distances.

Table 11 shows that after reducing contracting frictions in case CF3, the median increase in total sales is 0.20 percent, and the aggregate price index falls by 0.94 percent. The median change in intermediate sales and purchases increases by 0.48 percent and 0.20 percent, respectively. Moreover, the median change in the normalized number of buyers and sellers increases by 0.27 percent and 0.25 percent, respectively. Welfare increases by 0.95 percent, while average productivity falls by 0.12 percent. In Table 12, we show that in reducing contracting frictions, the least productive firms gain in terms of sales to the baseline, such that their weight in the aggregate increases. This drives the average productivity down.

5 Conclusions

We shed light on how cultural proximity shapes the formation of production networks and its implications for welfare. We first provide empirical evidence on the role of cultural proximity for inter-firm trade by leveraging a new dataset of firm-to-firm transactions from a large Indian state, along with data on firm owner names and their cultural proximity derived from India's caste and religious system.

We report three new stylized facts. First, culturally closer firms report higher sales between them, on the intensive margin. Second, culturally closer firms are more likely to trade with each other, on the extensive margin. Third, firms that are culturally closer report lower unit prices in their transactions.

We explore various mechanisms and find evidence most consistent with the importance of alleviating contracting frictions. Our effects are stronger for differentiated goods, which often rely on either formal or informal contract enforcement (Nunn, 2007; Rauch, 1999). We find that the importance of cultural proximity is elevated in regions with poor court quality (and so worse contract enforcement). Indeed, consistent with our narrative, the importance of court quality is only seen in trades of differentiated products, rather than homogeneous goods. Moreover, cultural proximity is more important for those firms that sell or buy more varieties, as they potentially face more contracting frictions. At the same time, culturally closer firms are less likely to cancel transactions between them. We understand these results as consistent with the possibility that cultural proximity is an informal mechanism that substitutes formal contract enforcement (Munshi, 2014, 2019).

We do not find sufficient evidence that hierarchies (and preference-based discrimination) matter, or that linguistic distance and the specialization in certain goods matter. Additionally, we do not find conclusive evidence that supports the idea that caste-based cartels play a role in our results. While family links, based on firm owners sharing the same surname, could be relevant, cultural proximity matters even over and above such connections.

We then build a quantitative general equilibrium model of firm-to-firm trade and cultural proximity. We introduce our measure of cultural proximity as a wedge that affects trade and matching costs, and estimate the key parameters of the model: the semi-elasticity of the trade cost to cultural proximity and the semi-elasticity of matching cost to cultural proximity. The model generates estimable specifications, that we take directly to the data. We use the estimated parameters to quantify the implications for welfare and other model-based statistics of implementing different policies. Welfare increases by 1.90 percent under social inclusion policies, falls by 0.93 percent under social isolation, and increases by 0.95 percent when reducing contracting frictions makes firms less reliant on cultural proximity.

In contexts like India, cultural and social networks may be used informally to overcome the lack of formal institutions that uphold contracts. Our paper is among the first to establish the consequences of these cultural ties in the context of trade. We study how social relationships influence firm-level decisions and quantify their importance for welfare. Our results have strong implications for policy. Promoting social inclusion and mixing via diversity-friendly policies can help facilitate matches and trade, with substantial implications for aggregate output and welfare. Furthermore, investing in reducing contracting frictions will allow firms to not have to rely on cultural ties, facilitating matches with more productive and low-cost suppliers, and once again improving economic well-being.

References

Alan, S., Baysan, C., Gumren, M., and Kubilay, E. (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics*, 136(4):2147–2194.

Alesina, A. and Giuliano, P. (2015). Culture and Institutions. *Journal of Economic Literature*, 53(4):898–944.

Alesina, A., Giuliano, P., and Reich, B. (2021). Nation-Building and Education. *The Economic Journal*, 131(638):2273–2303.

Alesina, A. F. and Reich, B. (2015). Nation Building. *Working Paper*.

Ali, S. N. and Miller, D. A. (2016). Ostracism and Forgiveness. *American Economic Review*, 106(8):2329–48.

Allen, T., Bilir, K., Chen, Z., and Tonetti, C. (2019). Knowledge Diffusion Through Networks. *Available at SSRN 3509835*.

Ambrus, A., Mobius, M., and Szeidl, A. (2014). Consumption Risk-Sharing in Social Networks. *American Economic Review*, 104(1):149–82.

Antras, P., Fort, T. C., and Tintelnot, F. (2017). The margins of global sourcing: Theory and evidence from US firms. *American Economic Review*, 107(9):2514–64.

Artiles, M. (2023). Within-Group Heterogeneity in a Multi-Ethnic Society. *Working Paper*.

Ash, E., Asher, S., Bhowmick, A., Bhupatiraju, S., Chen, D., Devi, T., Goessmann, C., Novosad, P., and Siddiqi, B. (2021). In-Group Bias in the Indian Judiciary: Evidence from 5 Million Criminal Cases. *Working Paper*.

Balmaceda, F. and Escobar, J. F. (2017). Trust in Cohesive Communities. *Journal of Economic Theory*, 170:289–318.

Bandyopadhyay, S., Coughlin, C. C., and Wall, H. J. (2008). Ethnic Networks and U.S. Exports. *Review of International Economics*, 16(1):199–213.

Banerjee, A. V. and Duflo, E. (2000). Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry. *The Quarterly Journal of Economics*, 115(3):989–1017.

Bau, N., Khanna, G., Low, C., and Voena, A. (2023). Traditional institutions in modern times: Dowries as pensions when sons migrate. Technical Report 31176, National Bureau of Economic Research.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2019). Unity in Diversity? How Intergroup Contact can Foster Nation Building. *American Economic Review*, 109(11):3978–4025.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K., and Moxnes, A. (2022). The Origins of Firm Heterogeneity: A Production Network Approach. *Journal of Political Economy*, 130(7):000–000.

Bhagavatula, S., Bhalla, M., Goel, M., and Vissa, B. (2018). Cultural Diversity on Corporate Boards and Firm Outcomes. *Working Paper*.

Bhalla, M., Chatterjee, I., Goel, M., and Khanna, G. (2024). Shared identity and entrepreneurship. Working Paper.

Bhattacharyya, A. (1943). On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109.

Bloch, F., Genicot, G., and Ray, D. (2008). Informal Insurance in Social Networks. *Journal of Economic Theory*, 143(1):36–58.

Board, S. (2011). Relational Contracts and the Value of Loyalty. *American Economic Review*, 101(7):3349–3367.

Boehm, J. and Oberfield, E. (2020). Misallocation in the Market for Inputs: Enforcement and the Organization of Production. *The Quarterly Journal of Economics*, 135(4):2007–2058.

Boehm, J., South, R., Oberfield, E., and Waseem, M. (2024). The Network Origins of Firm Dynamics: Contracting Frictions and Dynamism with Long-Term Relationships.

Boken, J., Gadenne, L., Nandi, T., and Santamaria, M. (2022). Community networks and trade.

Chen, M. X. and Wu, M. (2021). The Value of Reputation in Trade: Evidence from Alibaba. *The Review of Economics and Statistics*, 103(5):857–873.

Correia, S., Guimarães, P., and Zylkin, T. (2019). PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed Effects.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, Markups, and Trade Reform. *Econometrica*, 84(2):445–510.

De Sousa, J., Guillin, A., Lochard, J., and Silve, A. (2023). Trust and Specialization in Complexity: Evidence from US States. *Journal of Economic Behavior & Organization*, 216:332–353.

Depetris-Chauvin, E., Durante, R., and Campante, F. (2020). Building Nations through Shared Experiences: Evidence from African Football. *American Economic Review*, 110(5):1572–1602.

Faia, E., Mayer, M., and Pezone, V. (2021). The Value of Firm Networks: A Natural Experiment on Board Connections. *SAFE Working Paper*.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review*, 105(10):3150–82.

Fisman, R., Paravisini, D., and Vig, V. (2017). Cultural Proximity and Loan Outcomes. *American Economic Review*, 107(2):457–92.

Goraya, S. S. (2023). How Does Caste Affect Entrepreneurship? Birth versus Worth. *Journal of Monetary Economics*, 135:116–133.

Gradstein, M. and Justman, M. (2019). Cultural Interaction and Economic Development: An Overview. *European Journal of Political Economy*, 59:243–251.

Guiso, L., Sapienza, P., and Zingales, L. (2009). Cultural Biases in Economic Exchange? *The Quarterly Journal of Economics*, 124(3):1095–1131.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, pages 153–161.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics*, 123(2):441–487.

Iyer, L., Khanna, T., and Varshney, A. (2013). Caste and Entrepreneurship in India. *Economic and Political Weekly*, pages 52–60.

Khanna, G. (2020). Does Affirmative Action Incentivize Schooling? Evidence from India. *Review of Economics and Statistics*, 102(2):219–233.

Kone, Z. L., Liu, M. Y., Mattoo, A., Ozden, C., and Sharma, S. (2018). Internal Borders and Migration in India. *Journal of Economic Geography*, 18(4):729–759.

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies*, 74(3):791–819.

Macchiavello, R. and Morjaria, A. (2015). The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports. *American Economic Review*, 105(9):2911–45.

Munshi, K. (2014). Community Networks and the Process of Development. *Journal of Economic Perspectives*, 28(4):49–76.

Munshi, K. (2019). Caste and the Indian Economy. *Journal of Economic Literature*, 57(4):781–834.

Munshi, K. and Rosenzweig, M. (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review*, 106(1):46–98.

Nunn, N. (2007). Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics*, 122(2):569–600.

Rauch, J. E. (1996). Trade and Search: Social Capital, Sogo Shosha, and Spillovers. *National Bureau of Economic Research*.

Rauch, J. E. (1999). Networks Versus Markets in International Trade. *Journal of international Economics*, 48(1):7–35.

Rauch, J. E. and Casella, A. (2003). Overcoming Informational Barriers to International Resource Allocation: Prices and Ties. *The Economic Journal*, 113(484):21–42.

Rauch, J. E. and Trindade, V. (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130.

Richman, B. D. (2006). How Community Institutions Create Economic Advantage: Jewish Diamond Merchants in New York. *Law & Social Inquiry*, 31(2):383–420.

Ritzen, J., Easterly, W., and Woolcock, M. J. (2000). *On" Good" Politicians and" Bad" Policies: Social Cohesion, Institutions, and Growth*, volume 2448. World Bank Publications.

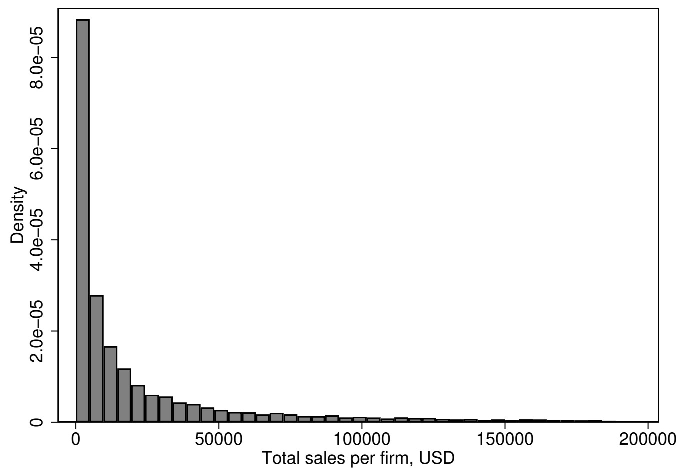
Schoar, A., Iyer, R. K., and Kumar, S. (2008). Importance of Ethnic Networks for Business Transactions of the Small Enterprises.

Startz, M. (2016). The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade. *Working Paper*.

Zhou, Y. (1996). Inter-Firm Linkages, Ethnic Networks, and Territorial Agglomeration: Chinese Computer Firms in Los Angeles. *Papers in Regional Science*, 75(3):265.

A Additional figures and tables

Figure A1: Distribution of total sales per firm



Notes: Figure shows distribution of total sales per firm up to the 90th percentile. Values converted to US dollars. Data corresponds to the year 2019.

Table A1: Data summary

Panel A: Aggregate statistics	
Number of firms	21,497
Number of sellers	10,010
Number of buyers	16,403
Transactions	823,923
Total sales	1.2 billion USD
Average sales per transaction	1,434 USD

Panel B: Statistics per firm						
	Average	Std. Dev.	Min.	Median	Max.	IQR
Sales per seller	118,069	698,015	0	7,972	28,300,000	35,411
Transactions per seller	82	834	1	6	60391	23
Matches per seller	4	11	1	2	322	3
Types of varieties sold per seller	3	6	1	1	100	2
Purchases per buyer	72,052	450,260	0	5,805	17,400,000	25,151
Transactions per buyer	50	276	1	7	16,059	26
Matches per buyer	3	4	1	1	147	2
Types of varieties bought per buyer	5	8	1	2	211	5

Notes: Sales and purchases converted to USD. Data corresponds to the year 2019.

B Robustness

B.1 Correction for selection bias

Helpman et al. (2008) suggest that the traditional gravity equation estimations, which we use for our intensive margin regressions, could be biased because of selection issues. Thus, in this section, we follow their proposed correction for selection bias.

In the first stage, we estimate the following linear probability model:

$$tr\left(\nu,\omega\right)=\iota_{\nu}+\iota_{\omega}+\delta BC\left(\nu,\omega\right)+\eta\Xi\left(\nu,\omega\right)+\xi B2B\left(\nu,\omega\right)+\epsilon\left(\nu,\omega\right), \tag{A1}$$

where we follow the nomenclature and the in-state sample from our extensive margin regressions. Here, we need an excluded instrument that affects only the extensive margin (i.e. the matching cost) and not the intensive margin (i.e. the trade cost). Thus, we consider the indicator variable $B2B\left(\nu,\omega\right)$ that equals 1 when both seller ν and buyer ω are in IndiaMART and equals 0 otherwise. As mentioned in Section 2.2, IndiaMART is the largest e-commerce platform for business-to-business (B2B) transactions in India.[29] Thus, the idea here is that it is easier for both firms to match if they take part in this platform.

We present the results of this first stage in Table A2. As before, the closer the firms are in cultural terms, the more likely it is they will trade. Additionally, if both firms participate in IndiaMART, the more likely the trade.

After the estimation, we recover the predicted probability of trading $\widehat{tr}\left(\nu,\omega\right)$, with which we calculate the latent variable

$$\widehat{\zeta}\left(\nu,\omega\right)=\Phi^{-1}\left(\widehat{tr}\left(\nu,\omega\right)\right),$$

where $\Phi^{-1}\left(\cdot\right)$ is the inverse CDF of the standard normal distribution. Following Heckman (1979), we obtain the inverse Mills ratio

$$\Upsilon\left(\widehat{\zeta}\right)=\frac{\phi\left(\widehat{\zeta}\left(\nu,\omega\right)\right)}{\Phi\left(\widehat{\zeta}\left(\nu,\omega\right)\right)},$$

where $\phi\left(\cdot\right)$ is the PDF of the standard normal distribution, and $\Phi\left(\cdot\right)$ is the CDF of the standard normal distribution.

[29] In 2019, there were between 5 and 6 million registered firms in IndiaMART (https://www.indiamart.com), which represented all firm size groups and all geographic regions in India.

For the second stage, we estimate

$$\ln y(\nu,\omega) = \iota_\nu + \iota_\omega + \delta BC(\nu,\omega) + \eta \Xi(\nu,\omega) + \upsilon \Upsilon\left(\widehat{\zeta}\right) + \epsilon(\nu,\omega), \tag{A2}$$

where $y(\nu,\omega)$ is the total positive sales of seller ν to buyer ω and the term $\Upsilon\left(\widehat{\zeta}\right)$ accounts for selection bias.

We present the second stage results in Columns 3 and 4 of Table A3. We must note that for computational reasons, we work with only an in-state sample, our results are not directly comparable to the baseline results from Table 1. Therefore, Columns 1 and 2 present the results with the in-state sample but without the correction for selection bias.

We find that not considering the correction for selection biases the coefficients downwards. This way, we conclude that the main results related to the intensive margin in the paper represent a lower bound of the effect of cultural proximity on trade.

Table A2: Correction for selection bias, first stage

Dep. Variable	Trade Indicator
BC	0.0010***
	(0.0001)
$B2B$	0.0016***
	(0.0003)
Obs.	5,628,290
Adj. R2	0.0106
FE	Seller, buyer, origin×dest.

Notes: Table shows the results of estimating Equation A1. The sample contains only in-state firms. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer, and it is equal to 0 otherwise. Data corresponds to the year 2019. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

Table A3: Correction for selection bias, second stage

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.148***	0.095***	0.223***	0.132**
	(0.035)	(0.029)	(0.074)	(0.055)
Inv. Mills rat.			0.503	0.246
			(0.421)	(0.298)
Obs.	26,238	26,238	26,238	26,238
Adj. R2	0.392	0.360	0.392	0.360
FE	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation A2. Columns 1 and 2 do not consider the correction for selection bias term. The sample contains only in-state firms. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

B.2 Alternative cultural proximity measure

In this section, we present an alternative measure of cultural proximity to that of the Bhattacharyya coefficient. Define the standard discrete distribution-based Kullback and Leibler (1951) divergence as

$$KL\left(\nu\|\omega\right)=\sum_{x=1}^{X}\rho_{\nu}\left(x\right)\log\left(\frac{\rho_{\nu}\left(x\right)}{\rho_{\omega}\left(x\right)}\right).$$

We have that $KL\left(\nu\|\omega\right)\geq 0$, where $KL\left(\nu\|\omega\right)=0$ when sellers and buyers have exactly equal probability distributions, while it will be higher the more different the two probability distributions are.[30] Intuitively, we can see this measure as the expected difference between two probability distributions. However, this proximity measure is not symmetric; that is, $KL\left(\nu\|\omega\right)\neq KL\left(\omega\|\nu\right)$. Consider our previous example where we record a transaction between a seller ν and a buyer with distribution ω, from which we calculate $KL\left(\nu\|\omega\right)$. If, in a second transaction, the roles of the firms revert, then the Kullback-Leibler divergence would be $KL\left(\omega\|\nu\right)$, implying the cultural proximity between the two firms has changed, when it should not change. To convert this measure into a symmetric one, we define

$$KL_{sym}\left(\nu\|\omega\right)=KL\left(\nu\|\omega\right)+KL\left(\omega\|\nu\right)=KL_{sym}\left(\omega\|\nu\right).$$

Notice this similarity measure needs $\rho_{\nu}\left(x\right)>0$ and $\rho_{\omega}\left(x\right)>0$ for all x. However, the probability of a firm belonging to a certain cultural group may be zero. In those cases, we replace that probability of zero for a probability $\varepsilon\to 0^{+}$ such that KL_{sym} is well-defined. Tables A4 and A5 show the regression results for the intensive margin, extensive margin, and unit prices. In this case, the higher the Kullback-Leibler divergence, the more culturally different the buyer from the seller. The results confirm the findings from the main text.

[30] This interpretation diverts from the standard use the Kullback-Leibler has in information theory, where a higher divergence means a higher information loss.

Table A4: Effect of cultural proximity on trade, intensive and extensive margins, Kullback-Leibler

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions	Trade Indicator	Trade Indicator
KL_{sym}	-0.004***	-0.003**	-0.005***	-0.003**	-0.00004***	-0.00004***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.00000)	(0.00000)
Log dist.	-0.023	-0.065***			0.00007	
	(0.015)	(0.011)			(0.00005)	
Obs.	32,678	32,678	32,843	32,843	5,606,627	5,628,290
Adj. R2	0.415	0.359	0.410	0.356	0.617	0.0106
FE	Seller, buyer	Seller, buyer	Seller, buyer, origin×dest.	Seller, buyer, origin×dest.	Seller, buyer	Seller, buyer, origin×dest.

Notes: Columns 1, 2, 3 and 4 show the results of estimating a modified version of Equation 1. Columns 5 and 6 show the results of estimating a modified version of Equation 2. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer in the year 2019, and it is equal to 0 otherwise. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. A higher Kullback-Leibler divergence means two firms are socially farther away. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

Table A5: Effect of cultural proximity on prices, Kullback-Leibler

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices	Log Prices
KL_{sym}	0.003**	0.003**	0.003**	0.002*	0.002**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log dist.	0.023	0.023	0.028*			
	(0.016)	(0.016)	(0.017)			
Obs.	230,744	230,744	226,645	235,001	236,617	230,900
Adj. R2	0.932	0.932	0.935	0.933	0.925	0.936
FE	Seller×HS, buyer	Seller×HS, buyer, month	Seller×HS, buyer, month×HS	Seller×HS, buyer, origin×dest.	Seller×HS, buyer, month, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating a modified version of Equation 3. Good g is defined according to 6-digit HS classification. Prices trimmed by 4-digit HS code at 5 and 95 percent. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Prices refer to unit prices of goods sold. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are multi-way clustered at the seller, 4-digit HS and origin-destination level. Standard errors are in parentheses. A higher Kullback-Leibler divergence means two firms are culturally farther away. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

B.3 Language

In this section, we check if the results we find are driven by language similarity. To do so, we follow the two language similarity measures from Kone et al. (2018). Define ϑ_i^l as the share of people with mother tongue l in district i. Then, the common language measure between districts i and j is

$$commlang_{ij} = \sum_l \vartheta_i^l \vartheta_j^l. \tag{A3}$$

We can also define a language overlap measure, defined as

$$overlang_{ij} = \sum_l \min\left\{\vartheta_i^l, \vartheta_j^l\right\}. \tag{A4}$$

In both cases, the larger the measures, the less likely it should be for people in these districts to face communication barriers. Then, we estimate the following intensive margin regression that considers language measures:

$$\ln y(\nu,\omega) = \iota_\nu + \iota_\omega + \delta BC(\nu,\omega) + \eta \Xi(\nu,\omega) + \zeta lang + \varepsilon(\nu,\omega), \tag{A5}$$

where $lang$ is either $commlang_{ij}$ or $overlang_{ij}$.

Table A6 presents the results. We find that none of the measures is statistically significant. This suggests that the cultural proximity result is not driven by firms sharing the same language.

Table A6: Effect of cultural proximity and language on trade, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.108***	0.068**	0.108***	0.068**
	(0.033)	(0.028)	(0.033)	(0.028)
commlang	-0.322	-0.126		
	(0.389)	(0.305)		
overlang			-0.419	-0.061
			(0.406)	(0.324)
Obs.	30,703	30,703	30,703	30,703
Adj. R2	0.409	0.357	0.409	0.357
FE	Seller, buyer	Seller, buyer	Seller, buyer	Seller, buyer

Notes: This table shows the results of estimating Equation A5. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. Standard errors two-way clustered at the seller and buyer level. Standard errors are in parentheses. *commlang* and *overlang* are language distance measures according to Equations A3 and A4.

B.4 Goods specialization

The cultural groups in India are, in many cases, defined by the production of specific goods (Munshi, 2019).[31] In this section, we study whether the cultural proximity results come from the fact that cultural groups specialize in the production and consumption of certain goods. If this were true, then the cultural proximity results would be due to different cultural groups specializing in certain goods, and therefore forming specific links with other cultural groups that either produce their inputs or consume their goods.

Here, we first must know which are the goods in which each cultural group specializes in selling and buying within our dataset. For this, we take the following steps:

1. We assign each firm to a unique cultural group. We do this by assigning each firm to the cultural group to which it has the highest probability of belonging according to its vector of probabilities $\boldsymbol{\rho}$.
2. We determine which is the most important 4-digit HS code in terms of sales and purchases for each cultural group.
3. We create a dummy variable $\mathbb{I}_g^{spec,seller}$ that equals 1 when a firm sells the good in which its cultural group specializes in selling. Similarly, we define dummy variable $\mathbb{I}_g^{spec,buyer}$ that equals 1 when a firm buys the good in which its cultural group specializes in buying.

Working with a version of our dataset at the transaction level, we estimate the regression

$$\ln n_g(\nu,\omega,t) = \iota_{\nu g} + \iota_{gt} + \iota_{\omega} + \delta BC(\nu,\omega) + \xi\left(BC(\nu,\omega) \times \mathbb{I}_g^{spec}\right) + \eta\Xi(\nu,\omega) + \epsilon_g(\nu,\omega)\,. \quad \text{(A6)}$$

Table A7 presents the results for the sales. First, if the cultural proximity results were only driven by cultural groups producing specific specialized goods, then we would expect the term on cultural proximity to be close to zero, and the term on the interactions to be statistically different from zero. However, we find that cultural proximity matters for all types of goods: for those in which a cultural group specializes and for those in which a cultural group does not specialize.

Second, in Column 2, we find that the coefficient on the interaction term is positive and statistically significant. Nevertheless, we lose this statistical significance after controlling for

[31] We can also understand this as certain cultural groups specializing in certain occupations.

additional variables in Column 4. This could point to cultural proximity mattering more for those goods in which cultural groups specialize in buying, but the result is not conclusive.

All in all, the results suggest that specialization does not play a role in the determination of the effect of cultural proximity on trade.

Table A7: Effect of cultural proximity on trade by good specialization, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Sales	Log Sales	Log Sales
BC	0.072***	0.071***	0.064***	0.064***
	(0.026)	(0.025)	(0.023)	(0.023)
$BC \times \mathbb{I}_g^{spec,seller}$	-0.016		0.135	
	(0.160)		(0.304)	
$BC \times \mathbb{I}_g^{spec,buyer}$		0.152***		0.185
		(0.008)		(0.118)
Obs.	226,039	226,039	229,719	229,719
Adj. R2	0.853	0.853	0.854	0.854
FE	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS	Seller×HS, buyer, month×HS, origin×dest.	Seller×HS, buyer, month×HS, origin×dest.

Notes: This table shows the results of estimating Equation A6. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the sales of specific good g for a given pair of seller and buyer. Good g is defined according to 6-digit HS classification. Sales were trimmed by 4-digit HS code at 5 and 95 percent. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors are two-way clustered at the seller and 4-digit HS level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer the two firms are. The number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019). $\mathbb{I}_g^{spec,seller}$ indicates the good g is the good in which the seller's cultural group specializes in selling. $\mathbb{I}_g^{spec,buyer}$ indicates the good g is the good in which the buyer's cultural group specializes in buying.

B.5 Industry pair linkages

In this section, we revise the intensive margin regressions after considering that there are pairs of industries that are bound to trade more than other pairs. For instance, perhaps certain castes happen to specialize in certain industries, and these industries are more likely to trade with each other. For this analysis, we add a selling firm's industry $\times$ buying firm's industry fixed effect to the intensive margin estimation. The sectors are based on the 4-digit HS code of the good with the highest sales for each firm. Therefore, we estimate

$$\ln y(\nu,\omega) = \iota_\nu + \iota_\omega + \delta BC(\nu,\omega) + \eta \Xi(\nu,\omega) + \left(\iota_\nu^{ind} \times \iota_\omega^{ind}\right) + \varepsilon(\nu,\omega), \tag{A7}$$

where ι_ν^{ind} and ι_ω^{ind} denote industry fixed effects for the seller ν and the buyer ω.

Table A8 presents the results. When compared to the results in Table 1, we find that, while the effect of cultural proximity is slightly higher, the main message prevails.

Table A8: Effect of cultural proximity after controlling for industries, intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.105**	0.089**	0.145***	0.104**
	(0.052)	(0.043)	(0.055)	(0.045)
Obs.	16,194	16,194	16,229	16,229
Adj. R2	0.414	0.326	0.395	0.308
FE	Seller, buyer, seller ind.×buyer ind.	Seller, buyer, seller ind.×buyer ind.	Seller, buyer, origin×dest., seller ind.×buyer ind.	Seller, buyer, origin×dest., seller ind.×buyer ind.

Notes: Columns 1, 2, 3 and 4 show the results of estimating Equation A7. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales and transactions refer to the total sales and total number of transactions for a given pair of seller and buyer, respectively. Data corresponds to the year 2019. Origin-destination fixed effect considers the district of the seller and the buyer. An industry is classified according to the 4-digit HS classification of the most sold good by each firm. Standard errors are two-way clustered at the seller and buyer level. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are. Number of observations varies between specifications due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

C Model derivations

C.1 Technology

A unique variety ω is produced by a single firm which minimizes its unit cost of production subject to its production technology, so

$$\begin{aligned}
\min_{\{m(\nu,\omega)\}} \quad & \int_{\nu\in\Omega(\omega)} m(\nu,\omega)\, p(\nu,\omega)\, d\nu + wl(\omega)\ , s.t. \\
& y(\omega) = \kappa_\alpha z(\omega)\, l(\omega)^\alpha\, m(\omega)^{1-\alpha}\ , \\
& m(\omega) = \left(\int_{\nu\in\Omega(\omega)} m(\nu,\omega)^{\frac{\sigma-1}{\sigma}}\, d\nu\right)^{\frac{\sigma}{\sigma-1}}\ , \\
& y(\omega) = 1.
\end{aligned}$$

Merge the first and third constraints, such that

$$\begin{aligned}
y(\omega) &= \kappa_\alpha z(\omega)\, l(\omega)^\alpha\, m(\omega)^{1-\alpha}\ , \\
1 &= \kappa_\alpha z(\omega)\, l(\omega)^\alpha\, m(\omega)^{1-\alpha}\ , \\
l(\omega)^\alpha &= \frac{1}{\kappa_\alpha z(\omega)\, m(\omega)^{1-\alpha}}, \\
&= \kappa_\alpha^{-1} z(\omega)^{-1}\, m(\omega)^{\alpha-1}\ , \\
l(\omega) &= \kappa_\alpha^{-\frac{1}{\alpha}} z(\omega)^{-\frac{1}{\alpha}}\, m(\omega)^{\frac{\alpha-1}{\alpha}}\ .
\end{aligned}$$

Rewrite the minimization problem, such that

$$\begin{aligned}
\min_{\{m(\nu,\omega)\}} & \int_{\nu\in\Omega(\omega)} m(\nu,\omega)\, p(\nu,\omega)\, d\nu + wl(\omega)\ , \\
& \int_{\nu\in\Omega(\omega)} m(\nu,\omega)\, p(\nu,\omega)\, d\nu + \kappa_\alpha^{-\frac{1}{\alpha}} wz(\omega)^{-\frac{1}{\alpha}}\, m(\omega)^{\frac{\alpha-1}{\alpha}}\ , \\
& \int_{\nu\in\Omega(\omega)} m(\nu,\omega)\, p(\nu,\omega)\, d\nu + \kappa_\alpha^{-\frac{1}{\alpha}} wz(\omega)^{-\frac{1}{\alpha}} \left(\int_{\nu\in\Omega(\omega)} m(\nu,\omega)^{\frac{\sigma-1}{\sigma}}\, d\nu\right)^{\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}}\ .
\end{aligned}$$

The first order condition with respect to $m(\nu,\omega)$ is

$$\begin{aligned}
0 &= p(\nu,\omega) + \kappa_\alpha^{-\frac{1}{\alpha}} w z(\omega)^{-\frac{1}{\alpha}} \left(\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}\right)(\dots)^{\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1}\left(\frac{\sigma-1}{\sigma}\right) m(\nu,\omega)^{\frac{\sigma-1}{\sigma}-1},\\
p(\nu,\omega) &= \kappa_\alpha^{-\frac{1}{\alpha}}\left(\frac{1-\alpha}{\alpha}\right) w z(\omega)^{-\frac{1}{\alpha}}(\dots)^{\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1} m(\nu,\omega)^{-\frac{1}{\sigma}},\\
m(\nu,\omega)^{\frac{1}{\sigma}} &= \frac{\kappa_\alpha^{-\frac{1}{\alpha}}\left(\frac{1-\alpha}{\alpha}\right) w z(\omega)^{-\frac{1}{\alpha}}(\dots)^{\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1}}{p(\nu,\omega)},\\
m(\nu,\omega) &= \frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}}\left(\frac{1-\alpha}{\alpha}\right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}}(\dots)^{\sigma\left(\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1\right)}}{p(\nu,\omega)^\sigma}.
\end{aligned}$$

Now, the first order condition with respect to $m(\nu,\omega)$ is

$$m(\nu,\omega) = \frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}}\left(\frac{1-\alpha}{\alpha}\right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}}(\dots)^{\sigma\left(\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1\right)}}{p(\nu',\omega)^\sigma}.$$

We divide both first order conditions, such that

$$\begin{aligned}
\frac{m(\nu,\omega)}{m(\nu',\omega)} &= \frac{\frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}}\left(\frac{1-\alpha}{\alpha}\right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}}(\dots)^{\sigma\left(\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1\right)}}{p(\nu,\omega)^\sigma}}{\frac{\kappa_\alpha^{-\frac{\sigma}{\alpha}}\left(\frac{1-\alpha}{\alpha}\right)^\sigma w^\sigma z(\omega)^{-\frac{\sigma}{\alpha}}(\dots)^{\sigma\left(\frac{\sigma}{\sigma-1}\frac{\alpha-1}{\alpha}-1\right)}}{p(\nu',\omega)^\sigma}},\\
&= \frac{\frac{z(\omega)^{-\frac{\sigma}{\alpha}}}{p(\nu,\omega)^\sigma}}{\frac{z(\omega)^{-\frac{\sigma}{\alpha}}}{p(\nu',\omega)^\sigma}},\\
&= \frac{p(\nu',\omega)^\sigma}{p(\nu,\omega)^\sigma},\\
m(\nu',\omega) &= \frac{p(\nu,\omega)^\sigma m(\nu,\omega)}{p(\nu',\omega)^\sigma}.
\end{aligned}$$

We plug this expression back into the expression for the composite of intermediates, so

$$\begin{aligned}
m(\omega) &= \left(\int_{\nu'\in\Omega(\omega)} m(\nu',\omega)^{\frac{\sigma-1}{\sigma}}\, d\nu\right)^{\frac{\sigma}{\sigma-1}}, \\
&= \left(\int_{\nu'\in\Omega(\omega)} \left(\frac{p(\nu,\omega)^{\sigma} m(\nu,\omega)}{p(\nu',\omega)^{\sigma}}\right)^{\frac{\sigma-1}{\sigma}} d\nu\right)^{\frac{\sigma}{\sigma-1}}, \\
&= p(\nu,\omega)^{\sigma} m(\nu,\omega) \underbrace{\left(\int_{\nu'\in\Omega(\omega)} p(\nu',\omega)^{1-\sigma}\, d\nu\right)^{\frac{\sigma}{\sigma-1}}}_{=\left(P(\omega)^{1-\sigma}\right)^{\frac{\sigma}{\sigma-1}}}, \\
&= p(\nu,\omega)^{\sigma} m(\nu,\omega) \left(P(\omega)^{1-\sigma}\right)^{\frac{\sigma}{\sigma-1}}, \\
&= p(\nu,\omega)^{\sigma} m(\nu,\omega)\, P(\omega)^{-\sigma}, \\
&= m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma}, \\
p(\nu,\omega)\, m(\nu,\omega) &= m(\omega)\, p(\nu,\omega)^{1-\sigma} P(\omega)^{\sigma}, \\
n(\nu,\omega) &= P(\omega)\, m(\omega)\, p(\nu,\omega)^{1-\sigma} P(\omega)^{\sigma-1}, \\
&= N(\omega)\, p(\nu,\omega)^{1-\sigma} P(\omega)^{\sigma-1},
\end{aligned}$$

which is the demand of firm ω from variety ν, where $P(\omega)^{1-\sigma} = \int_{\nu\in\Omega(\omega)} p(\nu,\omega)^{1-\sigma} d\nu$ is the price index faced by firm ω, $n(\nu,\omega) = p(\nu,\omega)\, m(\nu,\omega)$ is the expenditure of ω on variety ν, and $N(\omega) = P(\omega)\, m(\omega)$ is the total expenditure of firm ω.

The expression for the unit cost of production is

$$\begin{aligned}
c(\omega) &= \frac{w^{\alpha} P(\omega)^{1-\alpha}}{z(\omega)}, \\
&= \frac{P(\omega)^{1-\alpha}}{z(\omega)},
\end{aligned}$$

where wages $w = 1$ is the numeraire price.

Now, firms engage in monopolistic competition since they produce a unique variety. In particular, firm ν maximizes profits by selling its good to buyers ω subject to the demand for its intermediate, so

$$\begin{aligned}
\max_{\{p(\nu,\omega)\}} \quad & \int_{\omega\in\Omega(\nu)} \left(p(\nu,\omega) - d(\nu,\omega)\, c(\nu)\right) m(\nu,\omega)\,, s.t. \\
& m(\nu,\omega) = m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma},
\end{aligned}$$

where $d(\nu,\omega)$ is the iceberg cost of firm ν selling to ω. Rewrite the profit function $\pi(\nu,\omega)$, such that

$$\begin{aligned}
\pi(\nu,\omega) &= \left(p(\nu,\omega) - d(\nu,\omega)\, c(\nu)\right) m(\nu,\omega),\\
&= p(\nu,\omega)\, m(\nu,\omega) - d(\nu,\omega)\, c(\nu)\, m(\nu,\omega),\\
&= p(\nu,\omega)\, m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma} - d(\nu,\omega)\, c(\nu)\, m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma},\\
&= m(\omega)\, p(\nu,\omega)^{1-\sigma} P(\omega)^{\sigma} - d(\nu,\omega)\, c(\nu)\, m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma}.
\end{aligned}$$

The first order condition is

$$\begin{aligned}
[p(\nu,\omega)] : (1-\sigma)\, m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma}&\\
-(-\sigma)\, d(\nu,\omega)\, c(\nu)\, m(\omega)\, p(\nu,\omega)^{-\sigma-1} P(\omega)^{\sigma} &= 0,\\
(\sigma-1)\, m(\omega)\, p(\nu,\omega)^{-\sigma} P(\omega)^{\sigma} &= \sigma d(\nu,\omega)\, c(\nu)\, m(\omega)\, p(\nu,\omega)^{-\sigma-1} P(\omega)^{\sigma},\\
(\sigma-1) &= \sigma d(\nu,\omega)\, c(\nu)\, p(\nu,\omega)^{-1},\\
p(\nu,\omega) &= \left(\frac{\sigma}{\sigma-1}\right) c(\nu)\, d(\nu,\omega),\\
&= \mu c(\nu)\, d(\nu,\omega),
\end{aligned}$$

where $\mu = \frac{\sigma}{\sigma-1}$ is the markup. Finally, firm ω maximizes profits subject to the household demand for its good:

$$\begin{aligned}
\max_{\{q(\omega)\}} \left(q(\omega) - c(\omega)\right) y(\omega),\ s.t.&\\
y(\omega) = q(\omega)^{-\sigma} Q^{\sigma-1} Y.&
\end{aligned}$$

Rewriting profits:

$$\begin{aligned}
&= \left(q(\omega) - c(\omega)\right)\left(q(\omega)^{-\sigma} Q^{\sigma-1} Y\right),\\
&= q(\omega)^{1-\sigma} Q^{\sigma-1} Y - c(\omega)\, q(\omega)^{-\sigma} Q^{\sigma-1} Y.
\end{aligned}$$

Then, the first order condition is:

$$\begin{aligned}
[q(\omega)] : (1-\sigma)\, q(\omega)^{-\sigma} Q^{\sigma-1} Y - (-\sigma)\, c(\omega)\, q(\omega)^{-\sigma-1} Q^{\sigma-1} Y = 0, \\
(1-\sigma) - (-\sigma)\, c(\omega)\, q(\omega)^{-1} = 0, \\
q(\omega) = \left(\frac{\sigma}{\sigma-1}\right) c(\omega), \\
q(\omega) = \mu c(\omega),
\end{aligned}$$

where μ is the markup.

C.2 Preferences

A representative household maximizes its utility subject to its budget constraint, so

$$\max_{\{y(\omega)\}} \quad \left(\int_{\omega\in\Omega} y(\omega)^{\frac{\sigma-1}{\sigma}}\, d\omega\right)^{\frac{\sigma}{\sigma-1}}, \; s.t. \int_{\omega\in\Omega} q(\omega)\, y(\omega)\, d\omega \leq Y.$$

The first order condition with respect to firm ω is

$$\begin{aligned}
&[y(\omega)] : \left(\frac{\sigma}{\sigma-1}\right)(\dots)^{\frac{\sigma}{\sigma-1}-1}\left(\frac{\sigma-1}{\sigma}\right) y(\omega)^{\frac{\sigma-1}{\sigma}-1} = \lambda q(\omega), \\
&\lambda q(\omega) = (\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega)^{-\frac{1}{\sigma}},
\end{aligned}$$

where λ is the Lagrangian multiplier of the budget constraint, and $(\dots)$ is an aggregate term we do not write down since it will cancel out during the derivation. Now, the first order condition with respect to another firm ω' is

$$\lambda q(\omega') = (\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega')^{-\frac{1}{\sigma}}.$$

We then divide both first-order conditions, such that

$$\begin{aligned}
\frac{\lambda q(\omega)}{\lambda q(\omega')} &= \frac{(\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega)^{-\frac{1}{\sigma}}}{(\dots)^{\frac{\sigma}{\sigma-1}-1} y(\omega')^{-\frac{1}{\sigma}}}, \\
\frac{q(\omega)}{q(\omega')} &= \frac{y(\omega)^{-\frac{1}{\sigma}}}{y(\omega')^{-\frac{1}{\sigma}}}, \\
&= \frac{y(\omega')^{\frac{1}{\sigma}}}{y(\omega)^{\frac{1}{\sigma}}}, \\
y(\omega')^{\frac{1}{\sigma}} &= y(\omega)^{\frac{1}{\sigma}} \frac{q(\omega)}{q(\omega')}, \\
y(\omega') &= y(\omega)\left(\frac{q(\omega)}{q(\omega')}\right)^{\sigma}.
\end{aligned}$$

We plug this demand back into the budget constraint, which holds with equality, so

$$\begin{aligned}
Y &= \int_{\omega'\in\Omega} q\left(\omega'\right) y\left(\omega'\right) d\omega, \\
&= \int_{\omega'\in\Omega} q\left(\omega'\right) \left[y\left(\omega\right) \left(\frac{q\left(\omega\right)}{q\left(\omega'\right)} \right)^{\sigma} \right] d\omega, \\
&= y\left(\omega\right) q\left(\omega\right)^{\sigma} \underbrace{\int_{\omega'\in\Omega} q\left(\omega'\right)^{1-\sigma} d\omega,}_{=Q^{1-\sigma}} \\
&= y\left(\omega\right) q\left(\omega\right)^{\sigma} Q^{1-\sigma}, \\
&= \left(q\left(\omega\right) y\left(\omega\right)\right) q\left(\omega\right)^{\sigma-1} Q^{1-\sigma}, \\
&= x\left(\omega\right) q\left(\omega\right)^{\sigma-1} Q^{1-\sigma}, \\
x\left(\omega\right) &= q\left(\omega\right)^{1-\sigma} Q^{\sigma-1} Y,
\end{aligned}$$

which is the demand function for the unique variety of firm ω, where $Q^{1-\sigma} = \int_{\omega\in\Omega} q\left(\omega\right)^{1-\sigma} d\omega$ is a CES aggregate price index, and $x\left(\omega\right) = q\left(\omega\right) y\left(\omega\right)$ is the expenditure on a variety ω.

C.3 Gravity of intermediates

By plugging the pricing equation in the demand of firm ω for intermediates from firm ν, we derive the firm-level gravity equation

$$\begin{aligned}
n(\nu,\omega) &= p(\nu,\omega)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \\
&= (\mu c(\nu) d(\nu,\omega))^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \\
&= \mu^{1-\sigma} d(\nu,\omega)^{1-\sigma} c(\nu)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega), \\
\log(n(\nu,\omega)) &= \log\left(\mu^{1-\sigma} d(\nu,\omega)^{1-\sigma} c(\nu)^{1-\sigma} P(\omega)^{\sigma-1} N(\omega)\right), \\
&= \underbrace{\log\left(\mu^{1-\sigma}\right)}_{\iota} + \underbrace{\log\left(c(\nu)^{1-\sigma}\right)}_{\iota_\nu} + \underbrace{\log\left(P(\omega)^{\sigma-1} N(\omega)\right)}_{\iota_\omega} + \log\left(d(\nu,\omega)^{1-\sigma}\right), \\
&= \iota + \iota_\nu + \iota_\omega + (1-\sigma)\log(d(\nu,\omega)),
\end{aligned}$$

where ι is an intercept, ι_ν are seller fixed effects, and ι_ω are buyer fixed effects.

C.4 Equilibrium given the extensive margin

In this section, we derive the expression for the equilibrium objects given the structure of the production network. We first derive the recursive expression for prices, and then for total sales.

Recursive expression for prices. Consider the expression for the CES price index, so

$$
\begin{aligned}
P(\omega)^{1-\sigma} &= \int_{\nu\in\Omega(\omega)} p(\nu,\omega)^{1-\sigma}\, d\nu, \\
P(z')^{1-\sigma} &= \int p(z,z')^{1-\sigma}\, l(z,z')\, dG(z), \\
&= \int \left(\left(\frac{\sigma}{\sigma-1}\right) c(z)\, d(z,z')\right)^{1-\sigma} l(z,z')\, dG(z), \\
&= \mu^{1-\sigma} \int (c(z)\, d(z,z'))^{1-\sigma}\, dG(z), \\
&= \mu^{1-\sigma} \int \left(\frac{P(z)^{1-\alpha}}{z} d(z,z')\right)^{1-\sigma} l(z,z')\, dG(z), \\
&= \mu^{1-\sigma} \int \left(\frac{P(z)^{1-\alpha}}{z} d(z,z')\right)^{1-\sigma} l(z,z')\, dG(z), \\
&= \mu^{1-\sigma} \int P(z)^{(1-\alpha)(1-\sigma)}\, z^{\sigma-1} d(z,z')^{1-\sigma}\, l(z,z')\, dG(z).
\end{aligned}
$$

That is, the price index for firms of productivity z' can be expressed as a function of all other price indexes of firms z. This forms a system of equations we can solve.

Total sales. Consider the expression for total sales (i.e. sales to the household and firms), so

$$
\begin{aligned}
S(\nu) &= x(\nu) + \int_{\omega\in\Omega(\nu)} n(\nu,\omega)\, d\omega, \\
S(z) &= x(z) + \int n(z,z')\, l(z,z')\, dG(z'), \\
&= q(z)^{1-\sigma} Q^{\sigma-1} Y \\
&+ \int \left[\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} d(z,z')^{1-\sigma} c(z)^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z,z')\, dG(z'), \\
&= q(z)^{1-\sigma} Q^{\sigma-1} Y \\
&+ \int \left[\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} d(z,z')^{1-\sigma} \left[\frac{P(z)^{1-\alpha}}{z}\right]^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z,z')\, dG(z'), \\
&= q(z)^{1-\sigma} Q^{\sigma-1} Y \\
&+ \left[\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left(\frac{P(z)^{1-\alpha}}{z}\right)^{1-\sigma} \right] \int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z,z')\, dG(z'), \\
&= \frac{q(z)^{1-\sigma} Y}{Q^{1-\sigma}} \\
&+ \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} N(z') \right] l(z,z')\, dG(z'), \\
&= \frac{[\mu c(z)]^{1-\sigma} Y}{Q^{1-\sigma}} \\
&+ \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} \left(\frac{(1-\alpha) S(z')}{\mu}\right) \right] l(z,z')\, dG(z'), \\
&= \frac{\left[\mu \frac{P(z)^{1-\alpha}}{z}\right]^{1-\sigma} Y}{Q^{1-\sigma}} \\
&+ \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \left[\frac{1-\alpha}{\mu}\right] \int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} S(z') \right] l(z,z')\, dG(z'), \\
&= \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} \right] \\
&\left[\frac{Y}{Q^{1-\sigma}} + \left(\frac{1-\alpha}{\mu}\right) \left(\int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} S(z') \right] l(z,z')\, dG(z') \right) \right],
\end{aligned}
$$

where we use the fact that $N(z') = \frac{(1-\alpha)S(z')}{\mu}$. Given prices $P(z)$, this forms a system of equations for sales we can solve.

D Estimation and calibration

In this section we lay out the details on our model estimation and calibration strategy.

D.1 Estimation of key parameters

Cultural endowments $\boldsymbol{\rho}$. For the cultural endowments, we assume each firm ν has a probability vector $\boldsymbol{\rho_\nu} = [\rho_\nu(1), \ldots, \rho_\nu(452)]$ of belonging to each of the 452 cultural groups, we observe in the data. We further assume the elements of $\boldsymbol{\rho_\nu}$ are randomly drawn from a Dirichlet distribution, such that $\rho_\nu(1), \ldots, \rho_\nu(452) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_{452})$, where $\alpha_1, \ldots, \alpha_{452} > 0$ are concentration parameters.[32] The probability density for the Dirichlet distribution is

$$\rho_\nu(1), \ldots, \rho_\nu(452) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_{452}) = \frac{\Gamma\left(\sum_{x=1}^{452} \alpha_x\right)}{\prod_{x=1}^{452} \Gamma(\alpha_x)} \prod_{k=1}^{452} \rho_\nu(x)^{\alpha_x - 1},$$

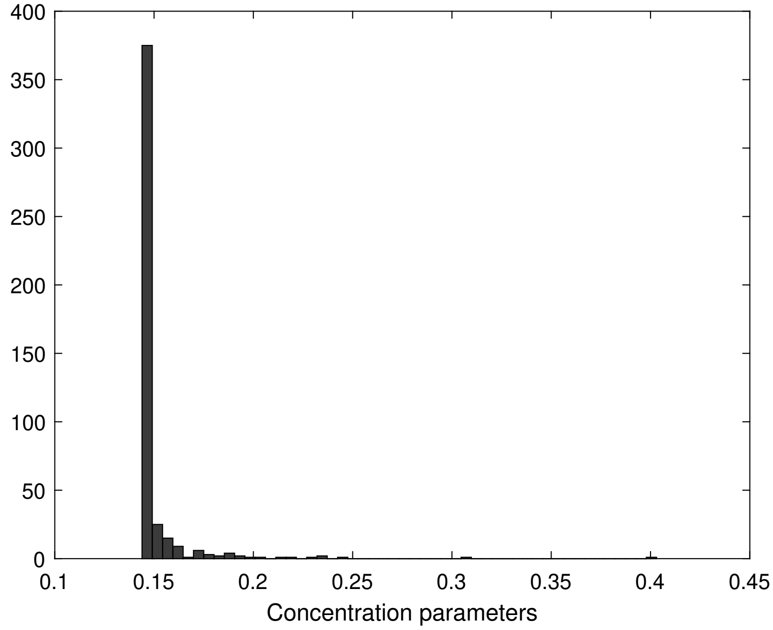
such that $\rho_\nu(x) \in [0,1]$, $\sum_{x=1}^{452} \rho_\nu(x) = 1$, where $\Gamma(.)$ is the gamma function and $\frac{\Gamma\left(\sum_{x=1}^{452} \alpha_x\right)}{\prod_{x=1}^{452} \Gamma(\alpha_x)}$ is a normalization constant. To ensure the theoretical Dirichlet distribution produces draws that are similar to the probabilities we see in the data, we estimate the vector $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1, & \ldots, & \alpha_{452} \end{bmatrix}$ parameters by maximum likelihood.[33] Let $\boldsymbol{\varrho} = \{\boldsymbol{\rho_1}, \ldots, \boldsymbol{\rho_N}\}$, where $\mathcal{N}$ is the total number of firms. Then, the log-likelihood function is

$$\ln pr(\boldsymbol{\varrho}|\boldsymbol{\alpha}) = \mathcal{N} \ln \Gamma\left(\sum_{x=1}^{452} \alpha_x\right) - \mathcal{N} \sum_{x=1}^{452} \ln \Gamma(\alpha_x) + \mathcal{N} \sum_{x=1}^{452} (\alpha_x - 1)\left(\frac{1}{\mathcal{N}} \sum_{\nu=1}^{\mathcal{N}} \ln \rho_\nu(x)\right). \quad \text{(A8)}$$

We present the estimated parameters in Figure A2.

[32] For a given x, the higher this parameter, the more disperse the realizations of $\rho_\nu(x)$ are across firms ν.
[33] For this, we use the Matlab toolboxes `fastfit` and `lightspeed` by Tom Minka.

Figure A2: Histogram of estimated concentration parameters for Dirichlet distribution



Notes: Estimated concentration parameters for a Dirichlet distribution according to the maximum likelihood estimation from Equation A8.

Trade costs d. From Equation 16 we need an estimate for β. Our setup produces a clear empirical counterpart that we already estimated in the reduced-form section, conditional on high-dimensional fixed effects. So we obtain an estimate for this parameter by linking the theoretical gravity Equation 17 to the empirical gravity equation results (Column 1 from Table 1). Thus, we obtain $\beta = -0.03$.[34]

Matching cost F. From Equation 13, we need an estimate for γ. We do this in two steps. First, using the extensive margin sample we run the following estimation

$$\ln\left[n\left(z, z^{'}\right)\right] = \iota_z + \iota_{z'} + \iota_{od} + \delta BC\left(z, z^{'}\right) + \varepsilon\left(z, z^{'}\right), \tag{A9}$$

where we apply the inverse hyperbolic sine transformation to the dependent variable, so as to not lose the cases in which there is zero trade. With this, we recover

$$\ln \widehat{\left[n\left(z, z'\right)\right]} = \widehat{\iota}_z + \widehat{\iota}_{z'} + \widehat{\iota}_{od} + \widehat{\delta} BC\left(z, z^{'}\right),$$

[34]Even though the wedge also appears in the price Equation 15 of the model, we do not estimate this equation to identify β. The reason is that the price equation is not an equilibrium equation, while the gravity equation is. Also, for our simulations we add a constant to the trade cost, such that the minimum trade cost is equal to 1. Therefore, in our simulations we have $d\left(\nu, \omega\right) = \exp\left(-\beta + \beta BC\left(\nu, \omega\right)\right)$.

where the hats denote estimated parameters and $\ln \widehat{[n(z,z')]}$ are the predicted sales. This variable predicts what would be the sales for a pair of seller and buyer even in the case they did not actually trade in the data. Second, we combine and rearrange Equations 13 and 14, such that

$$l\left(z,z^{'}\right)=\int 1\left[\ln\left(\epsilon\left(z,z^{'}\right)\right)<\ln \widehat{[n(z,z')]}-\gamma BC\left(z,z^{'}\right)+\chi\right]dH\left(\epsilon\left(z,z^{'}\right)\right), \quad \text{(A10)}$$

where we use the fact that $\pi\left(z,z^{'}\right)=\frac{n\left(z,z^{'}\right)}{\sigma}$ and replace $\ln\left[n\left(z,z^{'}\right)\right]$ by its estimated counterpart $\ln \widehat{[n(z,z')]}$, while χ is a constant term. We estimate this last equation with a probit (assuming $\epsilon\left(z,z^{'}\right)$ is log-normally distributed).

We present the results of the estimation of the first and second stages in Table A9. We find that $\gamma=-0.11$, and statistically different from 0.[35]

[35]For our simulations we add a constant to the matching cost, such that the minimum matching cost is equal to $1+\kappa$. Therefore, in our simulations we have $F\left(z,z^{'}\right)=\kappa+\exp\left(-\gamma+\gamma BC\left(z,z^{'}\right)\right)$.

Table A9: Estimation for matching cost

	(1)	(2)
	1st Stage	2nd Stage
Dep. Variable	Sales (Hyperbolic Inverse Sine)	Trade Indicator
BC	0.014***	0.109***
	(0.001)	(0.010)
$\ln\left[\widehat{n\left(z, z'\right)}\right]$		6.097***
		(0.038)
Obs.	5,628,290	5,628,290
Adj. R2	0.011	-
Pseudo R2	-	0.160
FE	Seller, buyer, origin $\times$ destination	-

Notes: Column 1 shows the results of estimating Equation A9. Column 2 shows the results of estimating Equation A10. We winsorize $\ln \widehat{n\left(z, z'\right)}$ at 1 percent and 99 percent. Sample only contains in-state firms. ***, ** and * indicate statistical significance at 1, 5, and 10 percent, respectively. Sales refer to the total sales for a given pair of seller and buyer in the year 2019. The trade indicator is equal to 1 if there was at least 1 recorded transaction for a given pair of seller and buyer in the year 2019, and it is equal to 0 otherwise. Standard errors are clustered at the seller and buyer level in Column 1. Standard errors are in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

D.2 Targeted and untargeted moments

Since the link function noise distribution affects how firms match between them, to identify the parameters related to this distribution we must target moments that are related to the extensive margin.

First, we target the mean of the log-outdegree $\ln\left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}}\right)$, where $\mathcal{N}_b\left(\nu\right)$ is the number of buyers a seller ν has. This moment captures the number of buyers a seller has on average, after normalizing by the total number of firms. We also target the mean of the log-indegree $\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)$, where $\mathcal{N}_s\left(\omega\right)$ is the number of sellers a buyer ω has. This moment captures the number of sellers a buyer has on average, after normalizing by the total number of firms. Because these two moments are related to magnitude of the matching, they should inform us about the mean of the link function noise distribution $\mu_{\ln(\epsilon)}$ and the scaling constant for the pairwise matching cost κ.

Second, this being mostly a seller-oriented model, to identify the standard deviation of the link function noise distribution $\sigma_{\ln(\epsilon)}$ we target the standard deviation of the log-outdegree $\ln\left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}}\right)$. Lastly, to identify the standard deviation of the log-productivity distribution, we must choose a moment that is related to the standard deviation of the intensive margin. Thus, we target the standard deviation of the log-normalized intermediate sales $\ln\left(\frac{\widetilde{N}(\nu)}{\mathcal{N}_b(\nu)}\right)$, where $\widetilde{N}(\nu)$ is the total intermediate sales a seller ν makes and we normalize by the number of buyers this seller has.

The first untargeted moment we consider is the standard deviation of the log-indegree $\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)$. The second untargeted moment we examine is the standard deviation of the log-normalized intermediate purchases $\ln\left(\frac{N(\omega)}{\mathcal{N}_s(\omega)}\right)$, where we normalize by the number of sellers a buyer has.

The exact definition of the targeted and untargeted moments, as well as the construction of their empirical counterparts are as follows:

Outdegree and indegree

Data. In our dataset, for each firm i, we calculate the number of firms it sold to and the number of firms it bought from. Then, to normalize this measure, we divide this number by the total number of firms in our sample. Thus, for a specific firm i, we can understand this measure as the share of firms this specific firm i is connected to, both as a buyer and a seller.

Model. For this part, we start with the link function matrix, where each element $l\left(z, z^{'}\right)$ represents the pairwise probability that seller z will match with buyer $z^{'}$. For each seller z, we take the average $l\left(z, z^{'}\right)$ across all the possible buyers. This represents the proportion of firms that seller z will match to the total number of firms. We follow a similar procedure to calculate the number of sellers each buyer $z^{'}$ has.

Normalized intermediate sales and purchases

Data. In our dataset, for each firm i, we calculate the total sales to other firms and the total purchases from other firms. In the case of the sellers, we normalize this measure by dividing the total sales of firm i by the total number of buyers this firm has. We follow a similar procedure with the buyers to calculate the normalized intermediate purchases.

Model. We use the intermediate sales matrix, where each element $n\left(z, z^{'}\right)$ represents the total sales of intermediate goods from seller z to buyer $z^{'}$. We sum all the sales for each seller z and divide this number by the number of buyers it has. Thus, we obtain the normalized intermediate sales for a given seller. For the normalized intermediate purchases, we follow a similar procedure with the buyers.

D.3 SMM estimation

We estimate the model using a simulated method of moments (SMM), where we choose the numerical values of (i) the standard deviation of the log-productivity distribution $\sigma_{\ln(z)}$ and (ii) the mean $\mu_{\ln(\epsilon)}$, (iii) the standard deviation $\sigma_{\ln(\epsilon)}$ of the link function noise distribution and (iv) the scaling constant for the pairwise matching cost κ, so as to match the targeted moments.

Given the other parameters discussed in Section 4.2, we perform the estimation 100 times. In each of the 100 iterations, there is a different draw of the productivity z_ν and the vector of probabilities of belonging to each caste $\boldsymbol{\rho_\nu}$ for each firm ν. Table A10 shows the results for the estimated parameters after 100 iterations.

Table A10: SMM estimated parameters

Parameter		Value
Std. dev. of the log-productivity distribution	$\sigma_{\ln(z)}$	0.66
		(0.01)
Mean of the link function noise distribution	$\mu_{\ln(\epsilon)}$	34.09
		(0.74)
Std. dev. of the link function noise distribution	$\sigma_{\ln(\epsilon)}$	9.53
		(0.01)
Scaling constant for the pairwise matching cost	κ	12.07
		(0.84)

Notes: Coefficients represent average after estimating the model 100 times. Standard errors after 100 estimations in parentheses.

Goodness of fit

Table A11 shows how the model-based moments fare against their empirical counterparts. When it comes to the targeted moments, the model can very closely replicate the empirical ones. For the untargeted moments, the model gets reasonably close to the data.

Table A11: Targeted and untargeted moments

Targeted Moments		
	Data	Model
Log-outdegree, average	-9.21	-9.37
Log-outdegree, std. dev.	0.98	0.91
Log-int. sales, std. dev.	1.30	1.31
Log-indegree, average	-9.37	-9.21
Untargeted Moments		
	Data	Model
Log-indegree, std. dev.	0.76	0.68
Log-int. purchases, std. dev.	1.29	1.01

Notes: Data moments correspond to the year 2019. Model moments represent average after estimating and simulating the model 100 times.