Cultural Proximity and Production Networks*

Brian Cevallos Fujiy — University of Michigan
Gaurav Khanna — UCSD
Hiroshi Toma — University of Michigan

First draft: April, 2022
This draft: May, 2023

Abstract

We examine how cultural proximity shapes production networks, and how it affects aggregate welfare and productivity. We combine a new dataset of firm-to-firm trade for a large Indian state with information on cultural proximity between firms derived from India's caste and religious classifications. We find that larger cultural proximity between a pair of firms reduces prices and fosters trade at both intensive and extensive margins. We argue that these results are driven by increasing trust between firms due to their cultural proximity, which in turn solves contracting frictions. Guided by these stylized facts, we propose a quantitative firm-level production network model, where cultural proximity influences trade and matching costs. We derive estimable equations from the model and estimate the model parameters leveraging variation in the cultural group composition of firm owners. We quantify the welfare and productivity consequences of implementing social inclusion policies that shape the formation of production networks. Our counterfactual exercises indicate that social inclusion policies can raise welfare by as much as 1.76%, while social isolation lowers welfare by 1.45%. Reducing contracting frictions increases welfare by 0.87% via the channel of trade becoming less reliant on cultural proximity.

Keywords: cultural proximity, production networks, firm-to-firm trade

JEL Codes: D51, F19, O17

*We thank Treb Allen, Mathias Bühler, 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 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, and WAITS Conference for their useful suggestions. We thank Manaswini Bhalla, Manisha Goel, and Manaswini Rao for sharing their datasets on castes and district courts. Eli Mogel and Tyler Spencer provided excellent research assistance. We also thank the 2021 Hellman Fellowship for their generous support for this project.
Cevallos Fujiy: bcfujiy@umich.edu, Khanna: gakhanna@ucsd.edu, Toma: htoma@umich.edu

1 Introduction

Non-economic forces, such as *culture*—, religion, language, values, etc.—drive economic outcomes. The role of culture on agent behavior has been well documented in entrepreneurship, loan access, labor markets, marriage, and international trade (Bandyopadhyay et al., 2008; Fisman et al., 2017; Goraya, 2022; Guiso et al., 2009; Hasanbasri, 2019; Macchiavello and Morjaria, 2015; Munshi and Rosenzweig, 2016; Rauch, 1996; Rauch and Casella, 2003; Rauch and Trindade, 2002; Schoar et al., 2008; Startz, 2016; Zhou, 1996). At the same time, recent evidence increasingly shows how inter-firm trade and production networks have important aggregate implications for economic development and welfare (Antras et al., 2017; Bernard et al., 2009, 2019; Bernard and Moxnes, 2018; Bernard et al., 2022; Dhyne et al., 2021; Eaton et al., 2011, 2016; Huneeus, 2018; Lim, 2018; Munshi and Rosenzweig, 2016; Oberfield, 2018; Taschereau-Dumouchel, 2019). Despite their parallel importance, the mechanisms by which cultural proximity shapes production networks and their aggregate implications remain less understood. Understanding how and why cultural proximity affects firm linkages and trade, potentially allows policy-makers to better leverage social inclusion programs and foster economic development. In this paper, we examine how cultural proximity determines connections and trade within production networks, and quantify the implications of cultural links for welfare and productivity.

We first provide empirical evidence on the role of cultural proximity in inter-firm trade and the formation of production networks. To do this, we leverage a unique dataset of firm-to-firm transactions 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 ever trade with each other. This means the higher the cultural proximity, the higher the trade on the extensive margin as well. Third, firms that are culturally further apart report higher 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 the importance of contract enforcement. First, we show suggestive evidence that the effect we find of cultural proximity on trade is driven by differentiated goods, which often rely on either formal or informal contract enforcement (Boehm, 2015; Boehm and Oberfield, 2020; Nunn, 2007; Rauch, 1999). Indeed, we find that differentiated goods, are more likely to be produced in and bought by firms that are located in districts with higher contract enforcement (as proxied by lower court delays). We understand these findings as evidence that cultural proximity relates to contract enforcement and trust (Munshi, 2019, 2014).[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.

We further find that the more varieties a firm sells or buys, the more the trade intensity is affected by social proximity. We posit that the larger the amount of different varieties a firm sells or buys, the more firms it has to negotiate with, which increases the contracting frictions it faces. Then, in order to minimize the contracting frictions they face, firms will rely more on trading with culturally closer firms they trust.

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

[1] Munshi (2019) uses survey data to show that Indians trust people from their caste. He also gives an example on how the Indian diamond industry relies on community networking because of the deficient contract enforcement.

is placed higher than the other based on the caste-based hierarchy, allowing us to test for preference-based discrimination across the social hierarchy. We do not find sufficient evidence that hierarchies (and preference-based discrimination) across social groups matter for our social proximity results. In other tests, we find our results are less likely to be driven by firms sharing the same language or specialization in the production of certain goods.

Encouraged by these stylized facts, we build a quantitative general equilibrium model of firm-to-firm trade and cultural proximity. Firms produce goods by combining labor and intermediate inputs in a CES fashion. Firms sell their goods to a household as final goods and to other firms as intermediates. Firms engage in monopolistic competition, charging a constant markup on top of their marginal costs. In our model, cultural proximity is a friction that influences both trade costs and fixed matching costs. These costs allows us to match our empirical findings, and are a simple way to quantify the aggregate implications of cultural proximity.

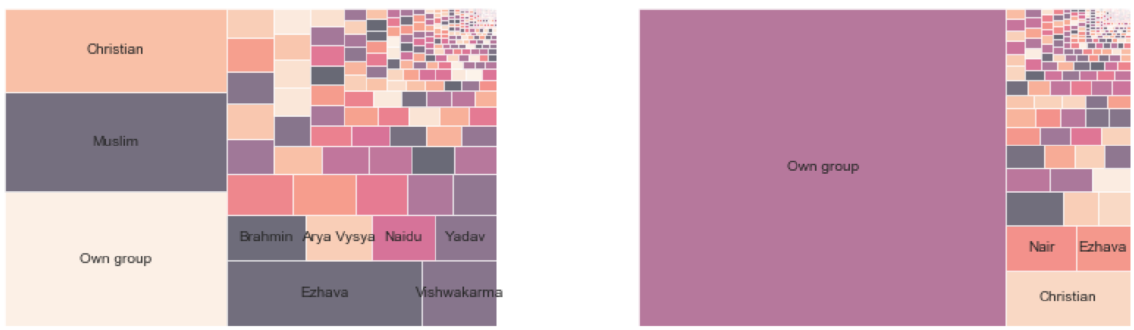
The model derives equations that precisely match their empirical counterparts in the previous section. We use these equations to estimate the key parameters of the model: the semi-elasticity of the trade cost to cultural proximity and the semi-elasticity of matching cost to cultural proximity. Our model 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 the closer two firms are in cultural terms, the lower the trade and matching costs are. Therefore, 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 1.76 percent under a diversity-friendly social inclusion policy. In contrast, welfare falls by 1.45 percent when we evaluate the effects of social isolation or exclusion. Finally, we show that policies that reduce contracting frictions raise welfare by 0.87 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: Figure shows the decomposition across buyers for the largest Hindu and non-Hindu cultural groups measured by probability-weighted sales. The Nair and Muslims accounted for 4.88 and 11.83 percent of total probability-weighted sales, respectively.

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.[2] In this case, cultural proximity naturally arises as a product of the inherent hierarchical structure of the caste system and the different 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

[2] In this paper, we consider the caste system and the religious groups as a proxy for cultural groups. There is a large historical legacy for the caste system to be considered as a device for discrimination, 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.

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.

This paper contributes to three strands of the literature. First, we contribute to the role of cultural proximity on economic outcomes such as trade (Bandyopadhyay et al., 2008; 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), entrepreneurship (Goraya, 2022), finance (Fisman et al., 2017), and labor markets (Hasanbasri, 2019; Munshi and Rosenzweig, 2016). We contribute by providing evidence and a framework to quantify the role of cultural proximity in production networks.[3] Second, it contributes to quantitative work on production networks (Antras et al., 2017; Bernard et al., 2009, 2014, 2019, 2022; Bernard and Moxnes, 2018; Dhyne et al., 2021; Eaton et al., 2011, 2016, 2022; Huneeus, 2018; Lim, 2018; Oberfield, 2018; Taschereau-Dumouchel, 2019). We contribute by embedding culture into a production network framework, and show how cultural proximity between firms can shape inter-firm trade, and what this implies for aggregate welfare. The uniqueness of our data in terms of measuring firm-to-firm transactions and the cultural group of owners, in combination with substantial variation across cultural groups, allow us to answer how cultural proximity shapes linkages and trade across the production network. Finally, 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 (Boehm and Oberfield, 2020; Boehm, 2015)[4] policies. We contribute by imple-

[3]In ongoing work, Boken et al. (2022) also show 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, 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 are also able to implement a Heckman selection bias correction model to estimate trade elasticities following Helpman et al. (2008).

[4]Contracting frictions can be either formal or informal. Boehm (2015); Boehm and Oberfield (2020) can simulate the effects of policies that improve court quality in India since formal contracting frictions are binding. We show that informal channels such as cultural proximity matter in the aggregate when implementing policies.

menting 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. In Section 4 we describe the model. In Section 5 we explain how we estimate the key parameters of the model. In Section 6 we analyze counterfactual scenarios. Section 7 concludes.

2 Background, data and construction of variables

2.1 Caste and Religion in India

India has a society that 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. According to this classification, people are classified across four possible groups called *Varnas*. From the most to the least privileged in hierarchical order, the four Varnas are *Brahmins*, *Kshatriyas*, *Vaishyas*, and *Shudras*. The Brahmins have historically enjoyed the most privileges, and are traditionally comprised of priests and teachers. The Kshatriyas are next in the hierarchy, usually associated with a lineage of warriors. The Vaishyas are third and are related to businessmen such as farmers, traders, among others. Finally, the Shudras are the most discriminated against and are the caste formed to be the labor class.

At the same time, Varnas are comprised by sub-groups called *Jatis* that were determined by factors such as occupation, geography, tribes, or language. In that sense, using Jatis as castes are appropriate for studying economic networks (Munshi, 2019), and from here on we use the notion of Jatis when referring to castes.

We also consider religious groups to define other cultural groups. The caste system is inherently based on Hindu religion, the predominant religion in India. While there are other religions in India which do not follow the caste system, they do relate to it: the other non-Hindu religions work as cultural groups of their own. We leverage information on firm owners belonging to both caste and religious groups to construct our measure of cultural proximity.

2.2 Data

Firm-to-firm trade. We leverage a firm-to-firm trade dataset for a large Indian state provided by the state's corresponding tax authority.[5] We use daily transactions 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 on 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 with the purpose of significantly increasing tax compliance.[6]

This data is provided by the tax authority of a large Indian state with a diversified production structure, 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 total value traded.

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

[5]While we use the term 'firm' in most parts of the paper, these data are actually at the more granular establishment level.

[6]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 of 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 (700 USD) must generate E-way bills. Transactions that have values lower than 700 USD can also be registered but it is not mandatory. There are three types of recorded transactions: (i) within-state trade, (ii) across-states trade, and (iii) international trade. For the purpose of this paper, we ignore international trade.

Firm owner names. The information about the name of the firm owners comes from two different sources. The first source 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 Indian state. Among these variables, we are provided with the name of the owner and/or of representatives of the firm.

To obtain firm-level characteristics of firms not registered in this state, we scrape the website *IndiaMART*,[7] the largest e-commerce platform for business-to-business (B2B) transactions in India. The website is comprised of 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.

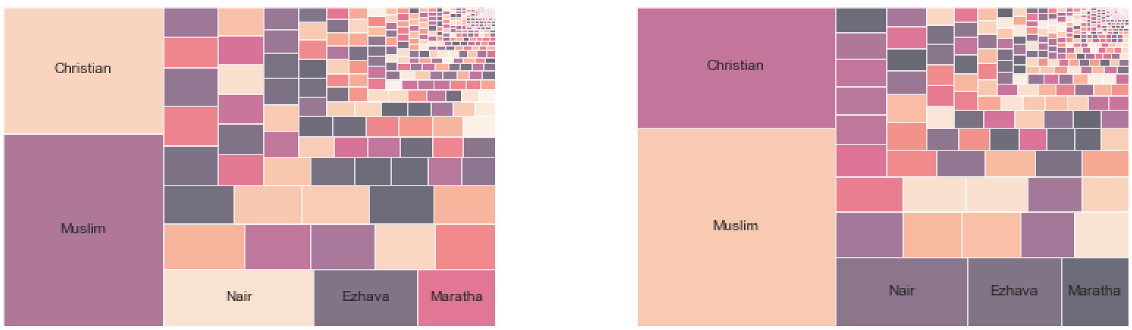
Matching owner names to cultural groups. We follow Bhagavatula et al. (2018) to match owner names to their Jatis (if the owners are of Hindu religion) or to their religion (in case the owners are not Hindu). Their procedure consists of using scraped data from Indian matrimonial websites that contain information on names, castes and religion. They

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

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 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.

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 rectangles reflects the share of sales and purchases.

Merged dataset. For the analytical part we merge the three previous datasets. We end up with a sample that contains information from 22,295 unique firms, of which there are 10,559 sellers and 16,980 buyers. In total, the sample comprises approximately 560 thousand transactions or 97 billion rupees (around 1.4 billion US dollars). We drop any registered transaction in which the seller and the buyer is the same parent firm. 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. 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

Firm-to-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. For obtaining the prices, we calculate the unit values. To do this, 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 of each good.

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 cultural group x. Based on these two distributions we construct the following measure of cultural proximity: the Bhattacharyya (1943) coefficient.

The Bhattacharyya (1943) coefficient between seller ν and buyer ω measures the level of overlapping between two different probability distributions.[8] 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 that of the buyer. In our context, this means the seller and the buyer have 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 exactly the same probability distribution of the buyer. This implies that the seller has the same probability of belonging to a group of certain cultural groups than the buyer or that their cultural proximity is the closest possible.[9] In robustness checks, we use the Kullback and Leibler (1951) divergence measure to measure cultural distance (Appendix C.1). All our results are qualitatively similar, and statistically significant when doing so.

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. The scatterplots show a higher Bhattacharyya

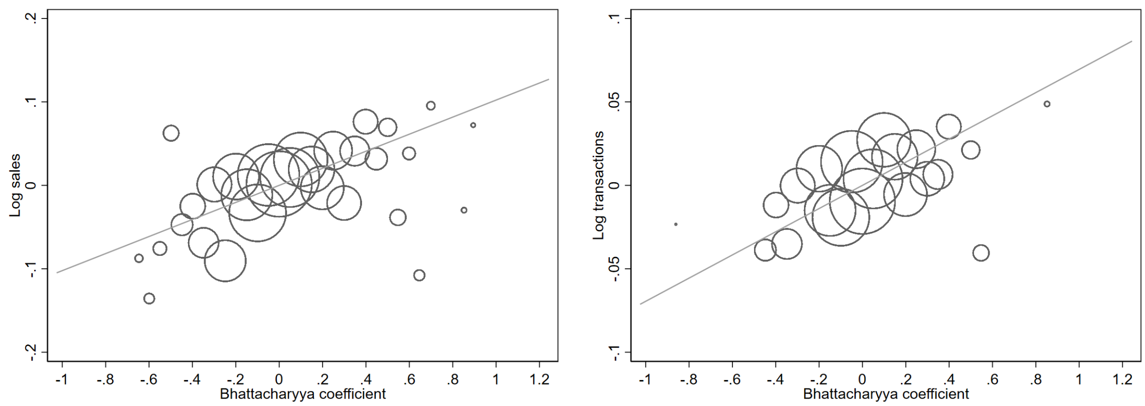
[8] Notice the Bhattacharyya coefficient is not the Bhattacharyya distance. The Bhattacharyya distance is defined as $BD(s, b) = -\log(BC(s, b))$. We prefer the Bhattacharyya coefficient because it is easier to interpret.

[9] 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. Our example shows the Bhattacharyya coefficient complies with this symmetry requirement, as $BC(\nu, \omega) = BC(\omega, \nu)$.

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

(a) Sales (b) # Transactions



Notes: Results residualized of seller fixed effects, buyer fixed effects and log distance. Equally distanced bins formed over the X axis. Size of bubbles represents number of transactions in each bin. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

Table 1: Effect of cultural proximity on trade, 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	0.100***	0.066**	0.129***	0.076***	0.0009***	0.0010***
	(0.033)	(0.027)	(0.034)	(0.028)	(0.0001)	(0.0001)
Log dist.	-0.023	-0.065***			0.0001	
	(0.015)	(0.011)			(0.0000)	
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 Equation (1). Columns 5 and 6 show the results of estimating Equation (2). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and buyer 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).

We now proceed to confirm the findings using a gravity equation. For transactions from firm

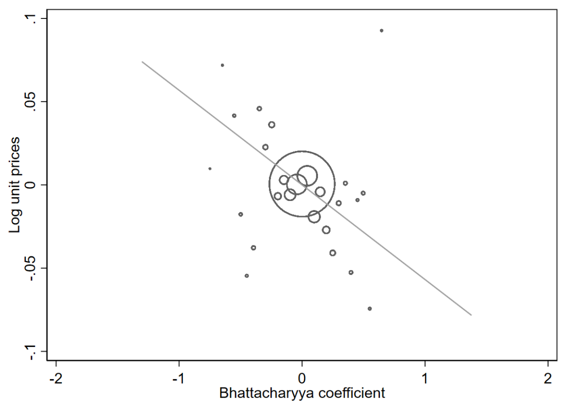
ν to firm ω in our sample we estimate

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

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, $dist(\nu,\omega)$ is the Euclidean distance between the pincodes in which the firms are located, ι_ν and ι_ω are seller and buyer fixed effects. Columns 1-4 of 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 firm when these firms are more alike in cultural terms. Columns 3 and 4 shows 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.

Fact 2: Cultural proximity increases the likelihood of ever 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. 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. This means, 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.

Figure 4: Effect of cultural proximity on prices



Notes: Results residualized of seller fixed effects and HS code fixed effects. Sectors defined according to 6-digit HS classification. Equally distanced bins formed over the X axis. Size of bubbles represents number of transactions in each bin. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

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, and 0 otherwise. With this variable we estimate a gravity-type specification:

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

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

Fact 3: Cultural proximity lowers prices. Figure 4 now uses buyer-seller-product groups and shows the residualized scatterplots between the similarity measures and the unit prices. We see 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 a seller-buyer-transaction-good version of our dataset and estimate

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

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, $\iota_{\nu \times g}$ is a seller-good fixed effect and $\iota_{g \times t}$ is a good-month fixed effect. We present the results in Table 2, which confirms the previous findings from the Figure: the closer the cultural proximity, the lower the unit value of the transactions.

Table 2: 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.

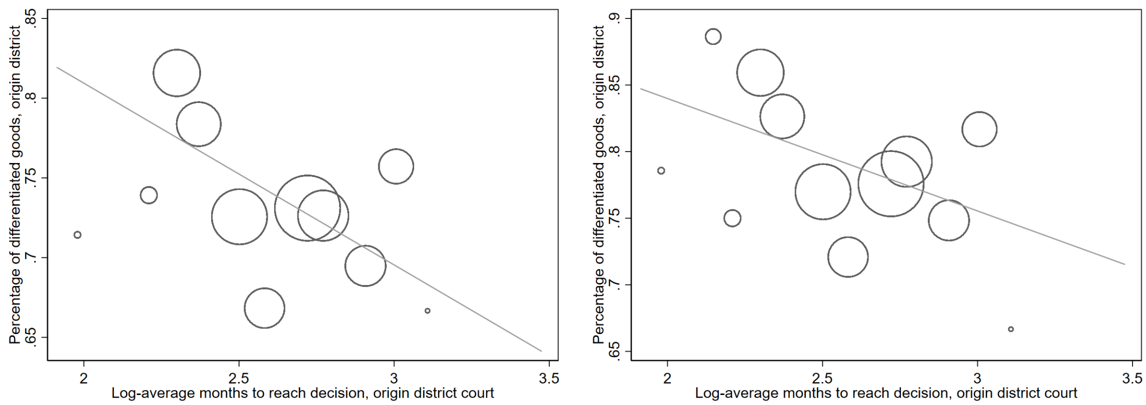
Notes: This table shows the results of estimating Equation (3). Good g is defined according to 6-digit HS classification. Prices trimmed by 4-digit HS code at 5 and 95 percent. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Origin-destination fixed effect 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 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).

3.1 Differentiated goods and court quality

To better understand the underlying forces driving these empirical patterns, we explore the importance of contract enforcement, and cultural hierarchies. First, in this section, we show evidence that suggests that the effect we find of cultural proximity on trade is driven by differentiated goods, which often rely on either formal or informal contract enforcement (Nunn, 2007). Then, we find that differentiated goods are more likely to be produced in and bought by firms that are located in districts with higher contract enforcement (as proxied by court delays). All in all, these analysis points that the stylized facts are likely driven by the desire of firms to reduce contracting frictions by trading with firms they trust. Here, cultural proximity arises as a proxy for knowing and trusting the other firm (Munshi, 2019, 2014).

Figure 5: Differentiated goods and court quality by district

(a) Origin district (b) Destination district



Notes: Scatter plot at the district level. Equally distanced bins formed over the X axis. Size of bubbles represents number of observations in each bin. The larger the log-average number of months for cases to reach a decision, the worse the district's court. Differentiated goods according to the conservative classification of Rauch (1999). The log-average number of months for cases to reach a decision comes from Ash et al. (2021), where for each district court in the 2010-2018 dataset we take into account the average months in between a case's date of filing and date of decision.

In order to bring in information on the type of product, we first disaggregate our data at the seller-buyer-transaction-good level. Then, we classify the goods into differentiated goods and non-differentiated goods based on the classification developed by Rauch (1999).[10] We estimate the following specification:

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

where $n_g(\nu,\omega,t)$ are the sales going from firm ν to firm ω of good g in month t and $\mathbb{I}_g^{diff}$ is an

[10] 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 in order to match to a suitable trade partner and will likely not abandon the commercial matches they have already made.

indicator for differentiated goods.[11] Table 3 presents the results for the sales. Our findings suggest that the baseline results of cultural proximity increasing trade are mostly driven by differentiated goods.

Table 3: 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 the 99, 95 and 90 percent level respectively. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. 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). $\mathbb{I}_g^{diff,con}$ 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).

What could be the reason behind differentiated goods driving the cultural proximity results? At the international trade level Nunn (2007) suggests contract enforcement is related the production of relationship-specific goods. To analyze this, we construct a measure of court quality at the district level.[12] Using data from Ash et al. (2021) we calculate the log-average number of months for cases to reach a decision in each district court between 2010 and 2018. The larger the log-average number of months for cases to reach a decision, the worse quality

[11] We use both the conservative and liberal classifications from Rauch (1999). The conservative classification 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.

[12] See Ash et al. (2021); Boehm and Oberfield (2020); Rao (2019) for references that analyze the effects of court quality in India.

this court has. Figure 5 shows that, in our dataset, districts with worse court quality sell and buy less differentiated goods, suggesting that differentiated products are more likely to be traded when contract enforcement is better.

Following an argument similar to Munshi (2019) and Nunn (2007), we interpret these findings as evidence that cultural proximity relates to contract enforcement and trust. 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. For buyers this could be not much of a hassle when it comes to homogeneous goods, as their suppliers are easily interchangeable. For sellers, this could be just a small problem as they can easily find other buyers. However, problems can arise if firms renege on their commitments related to differentiated goods. Suppliers and buyers in differentiated good markets are not easily replaceable. As a result suppliers of differentiated goods will only sell to buyers that they know and trust, while buyers of differentiated goods will do the same when choosing sellers.[13] Therefore, in these cases, trade will increase when firms trust and know each other, that is, when firms are culturally close.

3.2 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 generate indicators based on which is the Varna or religion for which a firm has the highest probability of belonging to.[14] We do not find evidence that hierarchies (and preference-based discrimination)

[13]We can relate our result to that of 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.

[14]While the Varna-based hierarchy only relates to the Hindu religion, we also place other religions in this hierarchy based on their income levels. We do this to prevent losing a large share of the sample in our

across social groups matter for our social proximity results.

We make use of two different indicators: $\mathbb{I}_{\nu_H \omega_L}$ and $\mathbb{I}_{\nu_L \omega_H}$. The first one indicates that the seller belongs to a higher hierarchy than the buyer. The second one indicates the buyer is placed below the seller in the social hierarchy. We include these two indicators by interacting them with our measure of cultural proximity. Table 4 presents the results for the intensive and extensive margins. The baseline category is that both firms belong to the same hierarchy. First place, we find the baseline coefficient is very similar to those of Table 1. Second, we find 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.

Table 4: 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 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 the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and buyer 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). 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.

estimations.

3.3 Additional specifications

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

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. Tables A3 and A4 show our baseline findings are robust to this alternative cultural proximity measure.

Language. We test whether the results we find are driven by language similarity. To do so, we follow the two linguistic distance measures from Kone et al. (2018). Table A5 shows that language does not affect the cultural proximity results already established.

Goods specialization. Cultural groups in India are, in many cases, defined by the production of specific goods (Munshi, 2019).[15] 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 Table A6 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.

Number of varieties sold and bought. We analyze whether firms that the social proximity results prevail for firms that sell and buy more varieties goods. To measure this, we count how many varieties of inputs a firm buys or how many varieties of goods a firm sells. In Table A7 we find the more varieties a firm sells or buys, the more the intensity of trade is affected by social proximity. Our interpretation of these findings is that the more varieties a firm sells or buys, the more contracting frictions it faces, caused by having to negotiate with either more suppliers or more clients. These firms, in order to minimize their

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

load of contracting frictions, will rely more on trading with counterparts in which they trust (i.e. firms that are culturally close).

Age of firms. We use the age of firms as a way for testing for taste-based discrimination. The idea is that, if there are taste-based preferences, then firms who sell to culturally close firms at lower prices are willing to forego profits because of their preferences (Becker, 1957). As a result, it should be more likely these firms will go bankrupt. We use the age of the firms as a proxy for which firms have not gone bankrupt. Tables A8 and A9 show the results the intensive margin regressions after controlling for age. If there was taste-based discrimination, then the interaction between the measure of cultural proximity and age should have a negative coefficient: Older firms should rely less on cultural proximity (i.e. they have been able to reach old age because of this). Nonetheless, we can only find weak evidence for taste-based discrimination. Thus, we are unable to establish that taste-based discrimination is the channel that drives our cultural proximity results.

Industry fixed effects. In the production matrix of an economy, there are sectors that are more likely to trade with others because of the nature of their activities. Then, in Table A10 we present the results for the intensive margin after adding a industry of seller $\times$ industry of buyer fixed effect. We find that the result of there being more trade between culturally close firm prevails.

Cancellations. Cultural proximity can also affect trade by making contract reneging less likely. A proxy for this in our dataset is the canceled transactions in between firms. In Table A11 we find that there is a lower probability of a transaction being canceled when the firms involved are close in cultural terms. While not conclusive, as we do not measure reneged contracts directly, this result suggests there could be a channel through which cultural proximity matters for having a lower rate of contract reneging.

Correction for selection bias. According to Helpman et al. (2008) the standard gravity equation estimations are biased as they do not account for selection issues. Therefore, we follow their suggested correction in Tables A12 and A13. 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). For this 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.

3.4 Discussion of stylized facts

The stylized facts show that a 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 we argue that contracting frictions could be the reason that drives the cultural proximity results. India is a country that suffers from 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 schemes or 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/or trust each other, and/or informally enforce contracts with social and reputational pressures. The higher the cultural proximity, the lower the contracting frictions.

Therefore, there would be more trade and lower prices, which is consistent with our previous findings. In Section 4 we present a simple theoretical framework in which cultural proximity affects contracting frictions and affects trade and prices. While our model is agnostic about why cultural proximity bridges the wedges in prices, the above discussion suggests that if contracting frictions drive initial trade barriers, then cultural proximity may reduce such frictions.

***Preference-based mechanisms* and discrimination.** We argue the results are unlikely to emerge from buyers having an inherent preference for buying from sellers culturally close to them. We could model this preference as 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 that is not consistent with our previous findings.

The stylized facts can arise from having sellers that show a preference for selling to buyers that are culturally close. It would imply the introduction of a supply shifter that is active for those buyers that are culturally close to the seller. However, we do not find conclusive evidence of this channel in Section C.5.

Discrimination from high-caste cultural groups against low-caste cultural groups may again reduce trade. Yet, in Section 3.2 we find this to be an unlikely driver of our empirical patterns. That is, 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.

4 Model

In this section we describe the model environment and define the equilibrium of the model. Appendix D contains further details.

4.1 Environment

Following Bernard et al. (2022), we build a quantitative firm-level production network model with heterogeneous firms and endogenous network formation. Firms maximize profits in two stages. First, each firm optimally determines its set of suppliers and buyers. Second, after the network is formed, firms maximize profits. We modify the original setting to make firms heterogeneous in productivity and their cultural endowments. We use these cultural endowments to construct a measure of cultural proximity between firms, which in turn influences trade costs and matching costs.[16]

Firms. There is a continuum of firms in the economy that operate under monopolistic competition and produce differentiated goods indexed by ω. We consider a roundabout production economy, so each firm produces by hiring labor from a representative household and by purchasing intermediate inputs from all the other firms in the economy.

The demand for firms come from other firms (intermediate inputs), and by a representative household (consumption goods).

Each firm has a technology

$$y\left(\omega\right) = \kappa_{\alpha} z\left(\omega\right) l\left(\omega\right)^{\alpha} m\left(\omega\right)^{1-\alpha}, \tag{5}$$

where $y\left(\omega\right)$ is output, $\kappa_{\alpha} \equiv \frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}$ is a normalization constant, $z\left(\omega\right)$ is firm-level productivity, $l\left(\omega\right)$ is labor, and $m\left(\omega\right)$ are intermediate inputs from other firms. In turn, the

[16] The main purpose of the model is to be used as an instrument to quantify the aggregate implications of cultural proximity between firms. Therefore, we model trade costs and matching costs in a reduced form fashion as a function of different factors, including cultural proximity. Nevertheless, there are many potential microfoundations for cultural proximity influencing trade costs based on hold up problems (Boehm, 2015; Boehm and Oberfield, 2020), reputation (Banerjee and Duflo, 2000; Chen and Wu, 2021), or loyalty (Board, 2011); and many potential microfoundations for cultural proximity influencing matching costs based on information and communication frictions (Ali and Miller, 2016; Allen et al., 2019; Balmaceda and Escobar, 2017), or risk-sharing (Ambrus et al., 2014; Bloch et al., 2008).

intermediate inputs are defined as a CES composite so

$$m(\omega) = \left(\int_{\nu \in \Omega(\omega)} m(\nu, \omega)^{\frac{\sigma-1}{\sigma}} \, d\nu \right)^{\frac{\sigma}{\sigma-1}},$$

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. By cost minimization we get

$$c(\omega) = \frac{P(\omega)^{1-\alpha}}{z(\omega)}, \tag{6}$$

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 markup pricing such that

$$p(\nu, \omega) = \mu c(\nu) \, d(\nu, \omega), \tag{7}$$

where $d(\nu, \omega) \geq 1$ is a pricing wedge that increases the price that seller ν charges to buyer ω, and $\mu \equiv \frac{\sigma}{\sigma-1}$ is the markup. We will define this wedge in the following paragraphs. We now derive the demand for intermediates, so

$$n(\nu, \omega) = p(\nu, \omega)^{1-\sigma} \, P(\omega)^{\sigma-1} \, N(\omega), \tag{8}$$

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 Equation (8) we can obtain the gravity equation as

$$\ln(n(\nu, \omega)) = \iota_\nu + \iota_\omega + (1-\sigma) \log(d(\nu, \omega)), \tag{9}$$

where ι_ν and ι_ω are seller and buyer fixed effects. This gravity equation relates directly to

Equation (1).

Pricing wedges. We assume the wedge is a function of different trade costs, including cultural proximity between firm owners due to ethnicity. Thus, we have

$$d(\nu,\omega) = \exp\left(\beta_1 dist(\nu,\omega) + \beta_2 BC(\nu,\omega) + \epsilon(\nu,\omega)\right), \tag{10}$$

where the parameters β_1 and β_2 are trade cost semi-elasticities. The wedge will be larger the longer the geographic distance and the lower the cultural proximity. From Equation (7) we have that the higher the cultural proximity, the lower the prices, which relates to stylized Fact 3. Likewise, from Equation (9) we have that the higher the cultural proximity, the higher the intermediate sales, which relates to stylized Fact 1.

Notice that the wedge encompasses different trade costs arising from geography, or institutional wedges arising from hold-up problems between sellers and buyers. The modeling of trade costs as icebergs as a function of geographic distance dates back to Samuelson (1954). Recent microfoundations of institutional wedges include Boehm and Oberfield (2020) and Boehm (2015). These microfoundations include informal channels to generate hold-up problems between firms such as the cultural proximity between them. Other potential microfoundations on the role of culture to determine trade are reputation (Banerjee and Duflo, 2000; Chen and Wu, 2021) and loyalty (Board, 2011).

Households. There is a representative household that demands goods from firms and inelastically supplies labor to them. To simplify, the representative household exhibits the same elasticity of substitution across goods σ as from firms. So, the representative household solves

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

where $P(\omega)$ is the price the household pays for good sold by ω, Ω is the set of firms in the economy, and Y is total income. This generates the demand for good ω

$$x(\omega) = P(\omega)^{1-\sigma} P^{\sigma-1} Y, \tag{11}$$

where $x(\omega) \equiv P(\omega) y(\omega)$ is the value of purchases from ω, and $P \equiv \left(\int_{\omega\in\Omega} P(\omega)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is a CES price index.

4.2 Equilibrium given production network

Here, we lay out the equilibrium conditions conditional on the structure of the network. Conditional on the formation of the network, firms only differ in productivity z, so we now identify each firm according to its productivity. Based on the price index of all of the goods acquired by firm $z^{'}$, we get

$$P\left(z^{'}\right)^{1-\sigma} = \mu^{1-\sigma} \int P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1} d\left(z, z^{'}\right)^{1-\sigma} l\left(z, z^{'}\right) dG(z), \tag{12}$$

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 household plus intermediates, so

$$S(z) = \left[\mu^{1-\sigma} P(z)^{(1-\alpha)(1-\sigma)} z^{\sigma-1}\right] \times \left[\frac{Y}{P^{1-\sigma}} D(z)^{1-\sigma} + \left(\frac{1-\alpha}{\mu}\right)\left(\int \left[d(z,z')^{1-\sigma} P(z')^{\sigma-1} S(z')\right] l(z,z')\, dG(z')\right)\right], \tag{13}$$

where $D(z) = \int_{\omega\in\Omega(\nu)} d(\nu,\omega)\, d\omega = \int d\left(z, z^{'}\right) l\left(z, z^{'}\right) dG\left(z^{'}\right)$ is the aggregated wedge for firm of productivity z.

4.3 Endogenous network

We endogeneize the formation of the production network by laying out the maximization problem of firms and how cultural proximity influences it. In particular, we allow for the cost of sellers and buyers matching to depend on their cultural proximity, which we can then estimate from the data. Before the formation of the network, firms are characterized by the tuple $\lambda = (z, \boldsymbol{\rho})$, where z is productivity, and $\boldsymbol{\rho}$ is the vector of probabilities of firm λ belonging to each cultural group. We can then construct a measure of cultural proximity according to the Bhattacharyya coefficient, such that

$$BC\left(z, z^{'}\right) = \sqrt{\sum_{x} \rho_{z}\left(x\right) \rho_{z^{'}}\left(x\right)}.$$

Now we describe how firms match. A seller z trades with a buyer $z^{'}$ only if it is profitable for the seller to do. To trade, the seller incurs in a pairwise matching cost $F\left(z, z^{'}\right)$.[17] Then, the share of seller-buyer pairs $\left(z, z^{'}\right)$ is

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

where $\pi\left(z, z^{'}\right)$ are the profits for seller z of selling to buyer $z^{'}$ and ϵ is an i.i.d. log-normal noise variable with mean 0 and standard deviation $\sigma_{\ln(\epsilon)}$. Intuitively, the link function can be understood as the probability a seller z will match to a seller $z^{'}$. We define the pairwise matching cost to be related to the cultural distance. Then

$$F\left(z, z^{'}\right) = \kappa + \exp\left(\gamma BC\left(z, z^{'}\right)\right), \quad (15)$$

where γ measures the sensitivity of the pairwise matching cost to the cultural distance and κ is a scaling constant.

[17] We assume that the matching cost is paid by the seller. For a further discussion on the importance of whether the seller or the buyer pays the fixed cost, see Huneeus (2018).

In the literature we find many potential ways of providing a microfoundation for cultural proximity affecting matching costs. These include information and communication frictions (Ali and Miller, 2016; Allen et al., 2019; Balmaceda and Escobar, 2017), risk-sharing (Ambrus et al., 2014; Bloch et al., 2008), among others.

From Equations (14) and (15), 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.

5 Estimation and calibration

Here we explain how we estimate the key parameters of the model on cultural endowments, (intensive) trade costs, and seller matching costs. We also describe how calibrate the remaining parameters of the model.

Cultural endowments $\boldsymbol{\rho}$. For the cultural endowments, we assume each firm ν has a probability vector $\boldsymbol{\rho_\nu} = [\rho_\nu(1), \dots, \rho_\nu(452)]$ 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 $\rho_\nu(1), \dots, \rho_\nu(452) \sim \mathcal{D}(\alpha_1, \dots, \alpha_{452})$, where $\alpha_1, \dots, \alpha_{452} > 0$ are concentration parameters.[18] The probability density function for the Dirichlet distribution is

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

such that $\rho_\nu(x) \in [0,1], \quad \sum_{x=1}^{452} \rho_\nu(x) = 1$, 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 $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1, & \dots, & \alpha_{452} \end{bmatrix}$

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

parameters by maximum likelihood.[19] 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

$$\ln pr\left(\boldsymbol{\varrho}|\boldsymbol{\alpha}\right) = \mathcal{N} \ln \Gamma\left(\sum_{x=1}^{452} \alpha_x\right) - \mathcal{N} \sum_{x=1}^{452} \ln \Gamma\left(\alpha_x\right) + \mathcal{N} \sum_{x=1}^{452} \left(\alpha_x - 1\right)\left(\frac{1}{\mathcal{N}} \sum_{\nu=1}^{\mathcal{N}} \ln \rho_\nu\left(x\right)\right). \tag{16}$$

Trade costs d**.** From Equation (10) we need an estimate for for $\{\beta_1, \beta_2\}$. We obtain estimates for these two parameters by linking the theoretical gravity equation (9) to the empirical gravity equation results (Column 1 from Table 3). Thus, we obtain $\{\beta_1, \beta_2\} = \{0, -0.03\}$.[20]

Matching cost F**.** From Equation (15), we need an estimate for γ. We do this in two steps. First, using the extensive margin sample we run the following estimation

$$\ln\left[n\left(z, z^{'}\right)\right] = \iota_z + \iota_{z^{'}} + \delta BC\left(z, z^{'}\right) + \gamma \ln\left(dist\left(z, z^{'}\right)\right) + \varepsilon\left(z, z^{'}\right), \tag{17}$$

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

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

where the hats denote estimated parameters and $\ln\left[\widehat{n\left(z, z'\right)}\right]$ 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 (14) and

[19] For this, we use the Matlab toolboxes `fastfit` and `lightspeed` by Tom Minka. We present the estimated parameters in Figure A1 in Appendix A.

[20] Even though the wedge also appears in the price equation (7) of the model, we do not estimate this equation to identify β_1 and β_2. 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 $d\left(\nu, \omega\right) = \exp\left(-\beta_2 + \beta_2 BC\left(\nu, \omega\right)\right)$.

(15), such that

$$l\left(z,z^{'}\right)=\int 1\left[\ln\left(\epsilon\left(z,z^{'}\right)\right)<\ln\widehat{\left[n\left(z,z'\right)\right]}-\ln\left(\sigma\right)-\gamma BC\left(z,z^{'}\right)\right]dH\left(\epsilon\left(z,z^{'}\right)\right), \quad (18)$$

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 $\ln\widehat{\left[n\left(z,z'\right)\right]}$.[21] We estimate this last equation with a probit regression (assuming $\epsilon\left(z,z^{'}\right)$is log-normally distributed). We find that $\gamma=-0.13$.[22]

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 $\sigma=3.94$. Following Bernard et al. (2022) 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 $\mathcal{N}=400$.

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). We explain this procedure below.

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

[21]For these estimations we ignore the scaling constant κ that appears in Equation (15).

[22]We present the results of the estimation in Table A1 in Appendix A. Also, for our simulations we add a constant to the matching cost, such that the minimum matching cost is equal to κ. Therefore, in our simulations we have $F\left(z,z^{'}\right)=\kappa+\exp\left(-\gamma+\gamma BC\left(z,z^{'}\right)\right)$.

must target moments that are related to the extensive margin.

First, we choose to target the mean of the log-normalized number of buyers $\ln\left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}}\right)$, where $\mathcal{N}_b(\nu)$ is the number of buyers a seller ν has; and the mean of the log-normalized number of sellers $\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)$, where $\mathcal{N}_s(\omega)$ is the number of sellers a buyer ω has. 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 variance of the log-normalized number of buyers $\ln\left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}}\right)$. Lastly, to identify the standard deviation of the log-productivity distribution, we must choose a moment that is related to the variance of the intensive margin. Thus, we target the variance 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.

The first untargeted moment we consider is the variance of the log-normalized number of sellers $\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)$. The second untargeted moment we examine is the variance of the log-normalized intermediate purchases $\ln\left(\frac{N(\omega)}{\mathcal{N}_s(\omega)}\right)$. The exact definition of the targeted and untargeted moments, as well as the construction of their empirical counterparts, appears in Appendix B.

Goodness of fit. After our matching procedure, we find the parameters $\sigma_{\ln(z)} = 0.88$, $\mu_{\ln(\epsilon)} = 64.30$,$\sigma_{\ln(\epsilon)} = 10.85$ and $\kappa = 14.80$. Table A2 in Appendix B 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.

6 Counterfactuals

We now present the results of various counterfactual exercises. First, we evaluate the effects of social mixing/inclusion and isolation policies, such that 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_2 and γ).

To evaluate each scenario, we measure what happens to various model-based statistics. Welfare is measured by real wage, $\mathcal{W} = \frac{w}{P}$. To quantify the impact on aggregate productivity, we consider a sales-weighted average productivity measure 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 the total economic activity, we measure total sales $\mathcal{S} = \sum_{\nu=1}^{N} S_\nu$, where S_ν are the total sales of firm ν. Additionally, we consider the average normalized intermediate sales $mean\left[\ln\left(\widetilde{N}\left(\nu\right)/\mathcal{N}_b\left(\nu\right)\right)\right]$, where $\widetilde{N}\left(\nu\right)$ are the total intermediate sales of seller ν, and the average normalized intermediate purchases $mean\left[\ln\left(N\left(\omega\right)/\mathcal{N}_s\left(\omega\right)\right)\right]$. For the prices, we compare the changes in the aggregate price index P. Finally, to study how matching between firms is affected, we present the results for the average normalized number of buyers, $mean\left[\ln\left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}}\right)\right]$, and the average normalized number of sellers, $mean\left[\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)\right]$.[23]

6.1 Social inclusion and social mixing policies

We analyze the effects of social inclusion or social 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

[23] In contrast to the previous sections, in this part we define the aggregate measures discretely. This is due to the simulations having a discrete number of firms, rather than a continuum.

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 actions programs may help incentivize students from different cultural groups to attend the same educative institutions. If these students then go on to become owners of the firms in the future, such policies may increase cultural proximity between these 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. 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.

Table 5 shows how the model statistics change in each counterfactual with respect to the baseline. In case CF1, we have that firms become the closest in cultural terms, so trade costs and matching costs go to their minimum possible. Aligned with our empirical facts, with lower trade costs, total sales increase by 2.76 percent, while the average intermediate sales and purchases go up by 1.52 percent and 1.15 percent, respectively. With the lower matching costs the average number of buyers grows by 1.07 percent, and the average number of sellers goes up by 1.00 percent. Also, because there are lower trade and matching costs, aggregate prices fall by 1.73 percent. With this, welfare increases by 1.76 percent. Besides welfare, another aggregate measure we analyze is average productivity, which falls by 0.13 percent. Yet, average productivity masks substantial compositional changes, as these results depend on whether the less productive firms are selling more or less with respect to the baseline case. We show in Table 6 that, in case CF1, when trade and matching costs decrease, the less productive firms match more and sell more, which increases their weight in the aggregate and lowers average productivity.

Table 5: Effect of cultural proximity on aggregate outcomes (counterfactual scenarios)

	CF1: Social inclusion/mixing	CF2: Social isolation	CF3: Reducing contracting frictions
Welfare	1.76	-1.45	0.87
Ave. productivity	-0.13	0.10	-0.06
Total sales	2.76	-2.23	1.37
Ave. normalized intermediate sales	1.52	-1.20	0.76
Ave. normalized intermediate purchases	1.15	-0.94	0.57
Ave. normalized number of buyers	1.07	-0.87	0.53
Ave. normalized number of sellers	1.00	-0.82	0.50
Agg. price index	-1.73	1.47	-0.87

Notes: We present the percentage gains or losses with respect to the baseline scenario. 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_2 and γ shrink by 50 percent.

Table 6: Change in sales by productivity quartiles

	CF1: Social inclusion/mixing	CF2: Social isolation	CF3: Reducing contracting frictions
1st quartile (most productive)	2.73	-2.21	1.35
2nd quartile	2.91	-2.35	1.44
3rd quartile	2.91	-2.31	1.44
4th quartile (least productive)	2.86	-2.32	1.42

Notes: We aggregate the sales of all firms that belong to a productivity quartile and calculate their percentage variation with respect to the baseline. 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_2 and γ shrink by 50 percent.

6.2 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.

When all firms are the furthest in cultural terms, trade costs and matching costs are the highest. Table 5 presents that in case CF2 total sales fall by 2.23 percent, average intermediate sales go down by 1.20 percent, average intermediate purchases fall by 0.94 percent and prices increase by 1.47 percent. There are also less matches, which is reflected by an average number of buyers that falls by 0.87 percentage points, and an average number of sellers that falls by 0.82 percentage points. As a result, welfare falls by 1.45 percent. Average productivity increases by 0.10 percent, relative to the baseline. Table 6 shows that in case CF2, every firm loses in terms of sales. However, the firms that lose the most are the least productive, which shrinks their weight in the aggregate and, thus, drives average productivity up.

6.3 Reducing contracting frictions

Now we turn to study which would be the effect of reducing contracting frictions. Following (Boehm, 2015; Boehm and Oberfield, 2020), 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_2 and γ shrink by 50 percent. This captures how reducing contracting frictions affect aggregate outcomes via the channel of trade becoming less reliant on cultural proximity.[24]

Table 5 shows that after reducing contracting frictions in case CF3 the total sales go up by 1.37 percent, average intermediate sales increase by 0.76 percent, average intermediate purchases grow by 0.57 percent and prices fall by 0.87 percent. The number of matches also increases, with the average number of buyers going up by 0.53 percent and the average number of sellers rising by 0.50 percent. Thus, welfare increases by 0.87 percent. Average productivity goes down by 0.06 percent. In Table 6 we show that in case CF3 all firms gain in terms of sales with respect to the baseline. Nonetheless, it is the lesser productive firms that gain the most, such that their weight in the aggregate increases. This drives the average productivity down.

7 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 and the formation of production networks, 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. That is, the higher the cultural proximity, the higher the trade in the intensive margin. Second, firms that are culturally closer are more likely to ever trade with each other. This means the higher the cultural proximity, the higher the trade in the extensive

[24] 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.

margin. Third, firms that are culturally further apart report higher unit prices in their transactions. We show evidence that suggests that the effect we find of cultural proximity on trade is stronger for differentiated goods, which often rely on either formal or informal contract enforcement (Nunn, 2007; Rauch, 1999). Indeed, we find that differentiated goods are more likely to be produced in and bought by firms that are located in districts with higher contract enforcement, as proxied by court delays. We understand these results as evidence that cultural proximity relates to contract enforcement and trust (Munshi, 2014, 2019).

We 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. We use the model and estimated parameters to quantify the implications for welfare and other model-based statistics of implementing different policies. Welfare increases by 1.76 percent when we evaluate a social inclusion policy, falls by 1.45 percent under social isolation and increases by 0.87 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 its 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, and so facilitate matches with more productive and low-cost suppliers, 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.

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.

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.

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 US 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.

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.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2009). The margins of US trade. *American Economic Review*, 99(2):487–93.

Bernard, A. B. and Moxnes, A. (2018). Networks and trade. *Annual Review of Economics*, 10:65–85.

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, 127(2):639–688.

Bernard, A. B., Van Beveren, I., and Vandenbussche, H. (2014). Multi-product exporters and the margins of trade. *The Japanese Economic Review*, 65(2):142–157.

Bhagavatula, S., Bhalla, M., Goel, M., and Vissa, B. (2018). Cultural diversity on corporate boards and firm outcomes.

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. (2015). The impact of contract enforcement costs onoutsourcing and aggregate productivity.

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.

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.

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.

Dhyne, E., Kikkawa, A. K., Mogstad, M., and Tintelnot, F. (2021). Trade and domestic production networks. *The Review of Economic Studies*, 88(2):643–668.

Eaton, J., Jinkins, D., Tybout, J., and Xu, D. (2016). Two-sided search in international markets. In *2016 Annual Meeting of the Society for Economic Dynamics*.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from French firms. *Econometrica*, 79(5):1453–1498.

Eaton, J., Kortum, S. S., and Kramarz, F. (2022). Firm-to-firm trade: Imports, exports, and the labor market.

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. (2022). How does caste affect entrepreneurship?: Birth vs worth.

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.

Hasanbasri, A. (2019). Explaining the female employment gap across religions and castes in India.

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.

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

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.

Lim, K. (2018). Endogenous production networks and the business cycle.

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.

Oberfield, E. (2018). A theory of input–output architecture. *Econometrica*, 86(2):559–589.

Rao, M. (2019). Judges, lenders, and the bottom line: Court-ing firm growth in india.

Rauch, J. E. (1996). Trade and search: Social capital, sogo shosha, and spillovers.

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.

Samuelson, P. A. (1954). The transfer problem and transport costs, ii: Analysis of effects of trade impediments. *The Economic Journal*, 64(254):264–289.

Schoar, A., Iyer, R. K., and Kumar, S. (2008). Importance of ethnic networks for business transactions of the small enterprises.

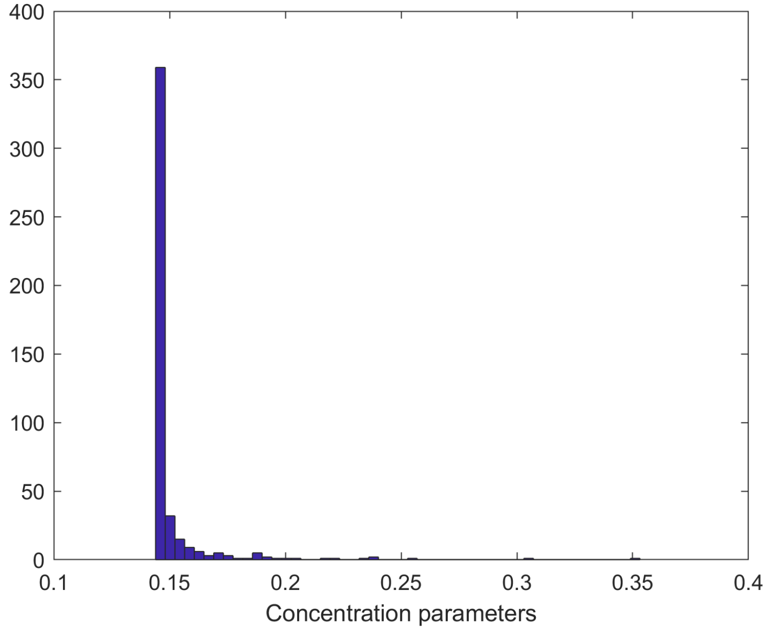
Startz, M. (2016). The value of face-to-face: Search and contracting problems in Nigerian trade.

Taschereau-Dumouchel, M. (2019). Cascades and fluctuations in an economy with an endogenous production network.

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: Histogram of estimated concentration parameters for Dirichlet distribution



Notes: Estimated concentration parameters for a Dirichlet distribution according to the maximum likelihood estimation from Equation (16).

Table A1: Estimation for matching cost

	(1)	(2)
	1st Stage	2nd Stage
Dep. Variable	Sales (Hyperbolic Inverse Sine)	Trade Indicator
BC	0.013***	0.131***
	(0.001)	(0.008)
$\text{ihs}\,\widehat{[n\,(z,z')]}$		8.340***
		(0.024)
Obs.	5,606,627	5,606,627
Adj. R2	0.595	-
Pseudo R2	-	0.453
FE	Seller, buyer	-

Notes: Column 1 shows the results of estimating Equation (17). Column 2 shows the results of estimating Equation (18). We winsorizeln $\widehat{n\,(z,z')}$ at 1 percent and 99 percent. Sample only contains in-state firms. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors clustered at the seller and buyer level in Column 1. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

B Targeted and untargeted moments

Normalized number of buyers and sellers

Data. In our dataset, for each firm i, we calculate the number 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 with respect to the total number of firms. We multiply this number by the total number of firms $\mathcal{N}$to obtain the number of buyers for each seller z. 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.

Table A2: Targeted and untargeted moments

Targeted Moments		
	Data	Model
$mean\left[\ln\left(\mathcal{N}_b(\nu)/\mathcal{N}\right)\right]$	-9.24	-9.48
$var\left[\ln\left(\mathcal{N}_b(\nu)/\mathcal{N}\right)\right]$	0.98	0.89
$var\left[\ln\left(\widetilde{N}(\nu)/\mathcal{N}_b(\nu)\right)\right]$	2.82	2.82
$mean\left[\ln\left(\mathcal{N}_s(\omega)/\mathcal{N}\right)\right]$	-9.39	-9.14
Untargeted Moments		
	Data	Model
$var\left[\ln\left(\mathcal{N}_s(\omega)/\mathcal{N}\right)\right]$	0.60	0.16
$var\left[\ln\left(N(\omega)/\mathcal{N}_s(\omega)\right)\right]$	2.73	0.56

Notes: The targeted moments are the mean of the log-normalized number of buyers $mean\left[\ln\left(\mathcal{N}_b(\nu)/\mathcal{N}\right)\right]$, the variance of the log-normalized number of buyers $var\left[\ln\left(\mathcal{N}_b(\nu)/\mathcal{N}\right)\right]$ and the variance of the log-normalized intermediate sales $var\left[\ln\left(\widetilde{N}(\nu)/\mathcal{N}_b(\nu)\right)\right]$, where $\widetilde{N}(\nu)$ are the total intermediate sales of seller ν. The untargeted moments are the mean of the log-normalized number of sellers $mean\left[\ln\left(\mathcal{N}_s(\omega)/\mathcal{N}\right)\right]$, the variance of the log-normalized number of sellers $var\left[\ln\left(\mathcal{N}_s(\omega)/\mathcal{N}\right)\right]$ and the variance of the log-normalized intermediate purchases $var\left[\ln\left(N(\omega)/\mathcal{N}_s(\omega)\right)\right]$.

C Additional specifications

C.1 Kullback-Leibler divergence

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.[25] 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, it is possible that the probability of a firm belonging to a certain cultural group is zero. In those cases we replace that probability of zero for a probability $\varepsilon\rightarrow 0^{+}$ such that KL_{sym} is well-defined. Tables A3 and A4 show the regression results for the intensive margin, unit prices and extensive margin, respectively. In this case, the higher the Kullback-Leibler divergence, the more culturally different the buyer from the seller. The results confirm the findings from

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

the main text.

Table A3: 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 the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and buyer level. Standard errors 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 A4: 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 the 99, 95 and 90 percent level respectively. 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 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).

C.2 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

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

We can also define a language overlap measure, defined as

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

In both cases, the larger the measures, the less likely it should be for people in these districts to face communication barriers. Table A5 presents the results of the intensive margin regression after considering the language measures. 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 A5: 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)
Log dist.	-0.025*	-0.065***	-0.029*	-0.065***
	(0.015)	(0.012)	(0.016)	(0.013)
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 a modified version of Equation (1). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors two-way clustered at the seller and buyer level. Standard errors in parentheses. A higher Kullback-Leibler divergence means two firms are socially farther away.

C.3 Goods specialization

The cultural groups in India are, in many cases, defined by the production of specific goods (Munshi, 2019).[26] In this section we study if the reason behind the cultural proximity results is actually cultural groups specializing in the production of certain goods and, given this, forming special bonds with their specific set of buyers.

First, we assign each firm to a unique cultural group. We do this by assigning each firm to the cultural group for which it has the highest probability of belonging to. In second place, we see which is the most important 4-digit HS code in terms of sales and purchases for each cultural group. We then match each firm to which is the good its cultural group specializes in selling and buying. Working with a version of our dataset at the seller-buyer-good level we run the regression

$$\ln n_g(\nu,\omega,t) = \iota_{\nu\times g} + \iota_{g\times t} + \iota_\omega + \delta BC(\nu,\omega) + \xi\left(BC(\nu,\omega)\times \mathbb{I}_g^{spec}\right) + \eta \ln dist(\nu,\omega) + \epsilon_g(\nu,\omega), \tag{A1}$$

where $\mathbb{I}_g^{spec}$ indicates if the good being traded is one in which either the cultural group of the selling firm specializes in selling or the cultural group of the buying firm specializes in buying. Table A6 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 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 too.

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

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 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 enough.

Table A6: 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 (A1). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. 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). $\mathbb{I}_g^{spec,seller}$ 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.

C.4 Number of varieties sold and bought

In this part we analyze whether firms that the cultural proximity results prevail for firms that sell and buy more varieties of goods. To measure this, we count how many 4-digit HS codes a firm buys or sells. Table A7 presents the results for the intensive margin, following a modified version of Equation (1). In our specifications $varieties_{\nu}^{sold}$ and $varieties_{\nu}^{bought}$ refer to the number of varieties sold and bought by the seller, while $varieties_{\omega}^{sold}$ and $varieties_{\omega}^{bought}$ refer to the number of varieties sold and bought by the buyer.

The results point to the effects of cultural proximity on trade being stronger when firms buy and sell more varieties. Our interpretation of these findings is that firms that buy and sell more varieties of goods have to face more contracting frictions, caused by having to negotiate more contracts. Then, these firms, in order to minimize their load of contracting frictions, will rely more on trading with counterparts in which they trust. Moreover, this explanation based on trust is compatible with the results related to differentiated goods from Section 3.1. In both cases we posit that the intensity of trade is driven by trust between firms, a coping mechanism to market imperfections in India.

Table A7: 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)
$BC \times varieties_{\nu}^{sold}$	0.089			
	(0.126)			
$BC \times varieties_{\nu}^{bought}$		0.121		
		(0.084)		
$BC \times varieties_{\omega}^{sold}$			0.112**	
			(0.051)	
$BC \times varieties_{\omega}^{bought}$				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)
$BC \times varieties_{\nu}^{sold}$	0.095			
	(0.105)			
$BC \times varieties_{\nu}^{bought}$		0.141**		
		(0.067)		
$BC \times varieties_{\omega}^{sold}$			0.104**	
			(0.042)	
$BC \times varieties_{\omega}^{bought}$				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 a modified version of Equation (1). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors 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.

C.5 Age of firms

In this section 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 who 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 establishment date from IndiaMART (the date in which a firm was established) and registration date from the tax authority (the data in 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 went eventually bankrupt, while the survivors were those firms that did not show these preferences.

Tables A8 and A9 show the results for a modified version of the intensive margin regressions according 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 being the reason behind our results.

Table A8: Effect of cultural proximity after controlling for establishment age of sellers (IndiaMART), intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.734**	0.489*	0.800**	0.479
	(0.355)	(0.296)	(0.371)	(0.311)
Log dist.	0.070*	0.002		
	(0.040)	(0.031)		
$BC\times$Log age seller	-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
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 a modified version of Equation (1). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and buyer 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). Age of seller comes from data reported in IndiaMART.

Table A9: Effect of cultural proximity after controlling for registration age of sellers (tax authority), intensive margin

	(1)	(2)	(3)	(4)
Dep. Variable	Log Sales	Log Transactions	Log Sales	Log Transactions
BC	0.164	0.217**	0.150	0.233**
	(0.115)	(0.094)	(0.116)	(0.097)
Log dist.	-0.044**	-0.075***		
	(0.019)	(0.014)		
$BC\times$Log age seller	-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 a modified version of Equation (1). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Standard errors two-way clustered at the seller and buyer 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). Age of seller comes from data reported by the tax authority.

C.6 Industry fixed effects

In this section revise the intensive margin regressions after considering the fact that there are pair of industries that are bound to trade more than other pairs. For this, we add a industry of seller $\times$ industry of buyer fixed effect to Equation (1). The sectors are based on the 4-digit HS code of the good with the highest sales for each firm. Table A10 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 A10: 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)
Log dist.	-0.065***	-0.094***		
	(0.022)	(0.018)		
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 (1). ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Origin-destination fixed effect considers the district of the seller and the buyer. Industry classified according to the 4-digit HS classification of the most sold good by each firm. Standard errors two-way clustered at the seller and buyer 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).

C.7 Cancellations

Among the diverse reasons for which cultural proximity could affect trade we can also consider reneged contracts, for which an available proxy in our dataset is the canceled transactions between a seller and a buyer. Thus, 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:

$$\mathbb{I}_{i,g}^{cancel}(\nu,\omega,t) = \iota_{\nu\times g} + \iota_{g\times t} + \iota_{\omega} + \delta BC(\nu,\omega) + \eta \ln dist(\nu,\omega) + \epsilon_{i,g}(\nu,\omega,t), \tag{A2}$$

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

Table A11 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. Therefore, our results are just suggestive of there being a channel whereby cultural proximity reflects in lower reneged contract, but are not conclusive.

Table A11: Cancellations

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable	Dummy Cancellation	Dummy Cancellation	Dummy Cancellation	Dummy Cancellation	Dummy Cancellation	Dummy Cancellation
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)
Log dist.	-0.003*	-0.003*	-0.003			
	(0.002)	(0.002)	(0.002)			
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 (A2) at the buyer-seller-transaction-good level. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Good g is defined according to 6-digit HS classification. Sales trimmed by 4-digit HS code at 5 and 95 percent. 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).

C.8 Correction for selection bias

Helpman et al. (2008) suggest that the traditional gravity equation estimations, which we use for our intensive margin regressions, are biased because for 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:

$$tr\left(\nu,\omega\right)=\iota_{\nu}+\iota_{\omega}+\delta BC\left(\nu,\omega\right)+\eta\ln dist\left(\nu,\omega\right)+\gamma B2B\left(\nu,\omega\right)+\epsilon\left(\nu,\omega\right), \tag{A3}$$

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 dummy 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. Thus, the idea here is that it is easier for both firms to match if they take part of this platform.

We present the results of this first stage in Table A12. As before, the closer the firms 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.

For the second stage we estimate

$$\ln y\left(\nu,\omega\right)=\iota_{\nu}+\iota_{\omega}+\delta BC\left(\nu,\omega\right)+\eta\ln dist\left(\nu,\omega\right)+\upsilon\Upsilon\left(\widehat{\zeta}\right)+\epsilon\left(\nu,\omega\right), \tag{A4}$$

where $y\left(\nu,\omega\right)$ 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 A13. We must note that because for this exercise we work with only an in-state sample, our results are not directly comparable to the baseline results from Table 1. Therefore, in Columns 1 and 2 we 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 A12: Correction for selection bias, first stage

Dep. Variable	Trade Dummy
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 (A3). Sample contains only in-state firms. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors two-way clustered at the seller and buyer level. Standard errors in parentheses. The higher the Bhattacharyya coefficient, the culturally closer two firms are.

Table A13: 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 (A4). Columns 1 and 2 do not consider the correction for selection bias term. Sample contains only in-state firms. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors two-way clustered at the seller and buyer 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).

D Model derivations

In this section we include details about the derivations of the theoretical model.

D.1 Firms

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

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

Merge the first and third constraints, such that

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

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

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

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

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

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 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.

D.2 Households

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

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

The first order condition with respect to firm ω is

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

where λ is the Lagrangian multiplier of the budget constraint, and $(\ldots)$ 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

$$\lambda P\left(\omega'\right) = \left(\ldots\right)^{\frac{\sigma}{\sigma-1}-1} y\left(\omega'\right)^{-\frac{1}{\sigma}}.$$

We then divide both first order conditions, such that

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

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

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

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

D.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

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

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

D.4 Equilibrium given network

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

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

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.