Cultural Proximity and Production Networks*

Brian Cevallos Fujiy
University of Michigan

Gaurav Khanna
University of California
San Diego

Hiroshi Toma
University of Michigan

[Preliminary and Incomplete]

September 17, 2022

Abstract

We examine how cultural proximity shapes production networks and its aggregate implications. 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 the intensive and extensive margins. Guided by these stylized facts, we propose a quantitative firm-level production network model, where cultural proximity influences the trade cost and the matching cost. 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. Reducing trade frictions across cultural groups boosts welfare by 1.61%.

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

JEL Codes: D51, F19, O17

*We thank Andrei Levchenko, Jagadeesh Sivadasan, Sebastian Sotelo, and seminar participants at University of Michigan and at the Ammersee Workshop on the Economics of Firms and Labor for their useful suggestions. We also thank Eli Mogel and Tyler Spencer for excellent research assistance.

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 agents' behavior has been documented in loan access, labor markets, marriage and international trade. At the same time, inter-firm trade and production networks have increasingly important aggregate implications for local economic development. Despite their parallel importance, the mechanisms by which cultural proximity shapes production networks and their aggregate implications remain less understood.

We examine how cultural proximity between firms shapes production networks and quantify its implications for welfare and productivity. We first provide empirical evidence on the role of cultural proximity for inter-firm trade and the formation of production networks. To do this, we leverage a new dataset of firm-to-firm transactions from a large Indian state, along with data on firms' owners names and their proximity arising from India's caste and religious system. For our purposes, castes and religious groups are useful as they provide a measurable notion of cultural groups.[1]

We report three new stylized facts. First, culturally closer firms report higher sales between them. This result holds after controlling for seller and buyer fixed effects, and geographic distance. Second, firms that are culturally further apart report higher prices in their transactions. In a context of low contract enforcement such as India, low cultural proximity works as a proxy for low trust in doing business, so a higher price prevails for a transaction to happen. Third, firms that are culturally closer are more likely to ever trade with each other.

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. We model cultural proximity as both a price wedge and a seller's cost to trade. Firms engage in monopolistic competition, so they charge a constant markup on top of their marginal costs. Importantly, price is also determined by a wedge, which is a function of the geographic distance and of how culturally similar the seller and the buyer are. At the same time, we also allow for cultural proximity to determine a seller's cost to trade with a buyer.

The model derives equations that exactly match their empirical counterparts in the previous section. We use these equations to estimate the key parameters of the model: the semi-

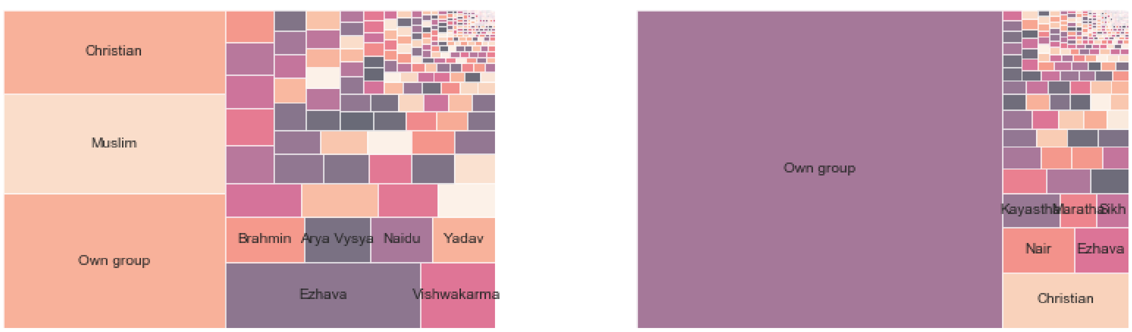
[1]We present the distribution of probability-weighted sales and purchases across cultural groups in Figure A1 in Appendix A.

elasticity of trade to cultural proximity and the semi-elasticity of the likelihood of trading to cultural proximity. Our model allows us to estimate both of these parameters externally. In line with our stylized facts, we find a positive semi-elasticity of both the intensive and extensive margin of trade to cultural proximity.

Finally, we use the model and estimated parameters to quantify the implications for welfare and productivity of implementing different inclusion policies that increase proximity between agents. First, we consider policies that increase the quality of contract enforcement, which is equivalent to an increase in the proximity term that determines prices. Second, we consider policies that improve the matching technology between sellers and buyers, which is equivalent to an increase in the proximity term that determines the matching cost. We quantify in scenarios that compare our baseline to both autarky (i.e. zero cultural proximity) and full trade (i.e. complete cultural proximity). We show that going from our baseline model to a full trade scenario increases welfare by 1.61% and average productivity slightly falls by 0.13%, whereas moving to autarky decreases welfare by 1.33%.

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 4.90 and 11.68 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, as a consequence, rely more on non-economic forces (Boehm and Oberfield, 2020; Munshi and Rosenzweig, 2016; Munshi, 2019). 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

[2]In this paper, we consider the caste system and the religious groups as a proxy for cultural groups.

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

This paper contributes to two strands of the literature. First, the paper contributes 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), finance (Fisman et al., 2017), and labor markets (Munshi and Rosenzweig, 2016; Hasanbasri, 2019). Second, it contributes to work on production networks (Antras et al., 2017; Bernard et al., 2009, 2014, 2019; Bernard and Moxnes, 2018; Bernard et al., 2022; Dhyne et al., 2021; Eaton et al., 2011, 2016, 2022; Huneeus, 2018; Lim, 2018; Oberfield, 2018; Taschereau-Dumouchel, 2019). We merge these two separate strands of the literature by providing both evidence and theory on how cultural proximity between firms can shape inter-firm trade, and what does that imply for aggregate welfare. The uniqueness of our data in terms of measuring firm-to-firm transactions, in combination with substantial variation across cultural groups, allow us to answer how cultural proximity shapes linkages and trade across the production network.

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 partially ruled 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

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.

economy. According to this classification, people are classified across four possible groups called *Varnas*. From the highest to the lowest in the hierarchical order, the four Varnas are *Brahmins*, *Kshatriyas*, *Vaishyas*, and *Shudras*. The Brahmins are at the top of the hierarchy which has been historically comprised by priests, and teachers. The Kshatriyas are the second ones on the hierarchy, usually associated with a lineage of warriors. The Vaishayas are third and they were related to businessmen such as farmers, traders, among others. Finally, the Shudras are the most discriminated against, and the caste was formed to be the labor class. There is an additional class called *Dalits*, also known as *Untouchables*; that is, people that were Dalits were subject to ostracism by the rest of society.

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 CEOs 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.[3] It contains information on daily transactions from January 2019 to December 2019, as long as any 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.[4]

This data is provided by the tax authority of a large Indian state with a diversified production structure, roughly 50\% urbanization rates, and high levels of population density. To

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

[4]For more details about the new E-Way bill system, see `https://docs.ewaybillgst.gov.in/`

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 on all the items contained within the transaction, the value of the transaction, the 6-digit HS code of the traded items and the quantity of each item and the units of the item. Since the data report both value and quantity of traded items, we construct unit values for each transaction. Each transaction also reports the pincode location of both selling and buying firms.[5] 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 CEO 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. That is, we cannot observe these characteristics for firms that did not formally register themselves in the state. Among these variables, we are provided with the name of the CEO and/or of representatives of the firm.

To obtain firm-level characteristics of firms not registered in this state, we scrape the website *IndiaMART*,[6] 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 CEO of the firm and the unique tax code identifier. Thus, we use the platform to obtain these variables for out-of-state firms. An example of how the website looks like for a given firm is in Figure A2 in Appendix A.

Matching CEO names to cultural groups. We follow Bhagavatula et al. (2018) to match CEO names to their Jatis (if the CEOs are of Hindu religion) or to their religion (in

[5] A pincode is the equivalent of the postal zip-code and determines postal code boundaries.
[6] `https://www.indiamart.com/`

case the CEOs are not Hindu). Their procedure consists of using scraped data from Indian matrimonial websites that contain information on names, castes and religion. They train a sorting algorithm that uses names as inputs and gives a probability distribution across cultural groups per name as outputs. We match these probability distributions to each CEO 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 CEOs trade with each other, they do not know each other's cultural group *ex ante*. Our sample finally consists of 452 cultural groups.

Merged dataset. For the analytical part we merge the three previous datasets. We end up with a sample that contains information from 22,437 unique firms, of which there are 10,564 sellers and 16,990 buyers. In total, the sample comprises 154 thousand transactions or 370 billion rupees (approximately 5 billion US dollars). We drop any registered transaction in which the seller and the buyer is the same firm. Each firm is linked to a unique pincode. Finally, we assign a sector to each firm based on the HS codes of the goods sold.

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, 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.[7] 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 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.[8] In robustness checks, we use the Kullback and Leibler (1951) divergence measure to measure cultural distance (Appendix C.1). Our results are qualitatively similar, and statistically significant when doing so.

[7] 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.

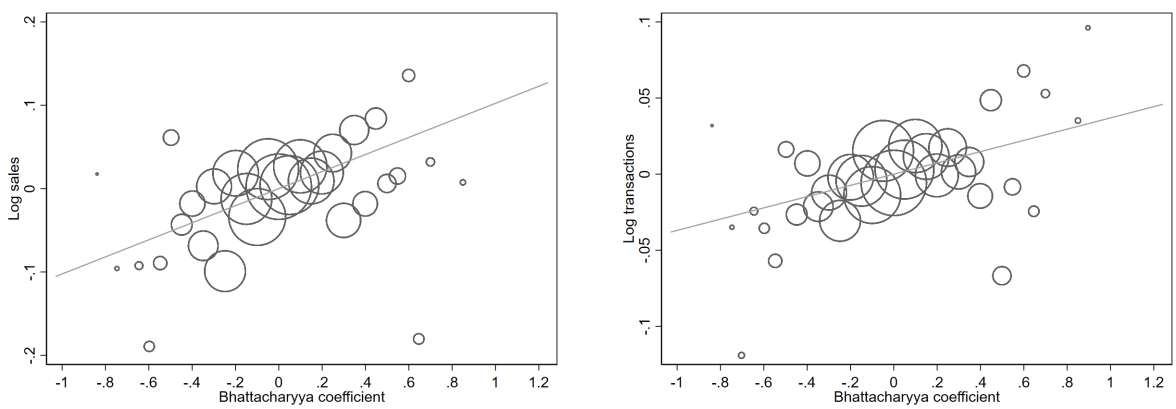
[8] For our purposes, it is important that the cultural proximity measure we use is symmetric. To see why, consider an example where, in our dataset, we have a transaction between a seller ν and a buyer ω, from which we obtain $BC(\nu, \omega)$. Further assume that in our dataset we record a second transaction 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 $BC(\nu, \omega) = BC(\omega, \nu)$.

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 2 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 coefficient (buyer and seller are more alike in their cultural group probability distributions) is related to a higher amount of sales and transactions.

Figure 2: 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 more culturally close two firms are.

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

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

where $y\left(\nu,\omega\right)$ is either the total sales $n\left(\nu,\omega\right)$ or total transactions $t\left(\nu,\omega\right)$ from seller ν to buyer ω, $BC\left(\nu,\omega\right)$ is the Bhattacharyya coefficient, $dist\left(\nu,\omega\right)$ is the Euclidean distance between the pincodes in which the firms are located, ι_{ν} and ι_{ω} are seller and buyer fixed effects. Table 1 presents the results of the intensive margin estimation, which confirm the preliminary findings from Figure 2. 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.

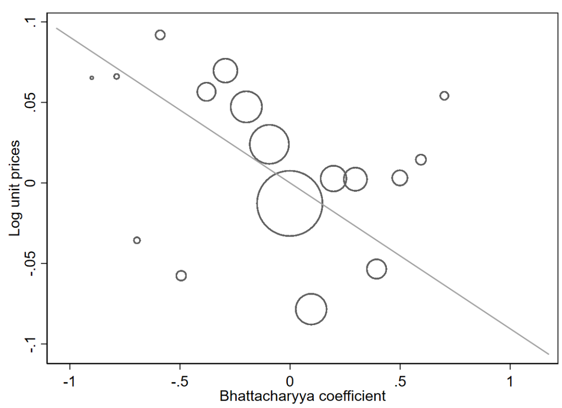
Fact 2: Cultural proximity lowers prices. Figure 3 now uses buyer-seller-product groups and shows the residualized scatterplots between the similarity measures and the unit prices. We see the higher the Bhattacharyya coefficient between two firms involved in a transaction, the lower the price that will be charged.

Table 1: Estimation, intensive margin

	(1)	(2)
Dep. Variable	Log Sales	Log Transactions
BC	0.097***	0.033*
	(0.036)	(0.018)
Log dist.	-0.001	-0.033***
	(0.017)	(0.007)
Constant	13.262***	1.194***
	(0.192)	(0.084)
Obs.	32,773	32,773
Adj. R2	0.439	0.309

Notes: This table shows the results of estimating Equation 1. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Regressions consider seller fixed effects and buyer fixed effects. Standard errors are clustered to the seller and buyer level. The higher the Bhattacharyya coefficient, the more culturally close two firms are.

Figure 3: 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 more culturally close two firms are.

To confirm the results, we estimate

$$\ln p_g(\nu,\omega) = \iota_{\nu\times g} + \iota_\omega + \delta BC(\nu,\omega) + \eta \ln dist(\nu,\omega) + \epsilon_g(\nu,\omega), \tag{2}$$

where $p_g(\nu,\omega)$ is the unit value of good g (at the 6-digit HS classification) sold by firm ν to firm ω and $\iota_{\nu\times g}$ is a seller-good fixed effect. We present the results in Table 2, which confirms the previous findings from the figures.

Table 2: Estimation, prices

	(1)	(2)	(3)
Dep. Variable	Log Prices	Log Prices	Log Prices
BC	-0.084***	-0.073**	-0.052*
	(0.031)	(0.029)	(0.030)
Log dist.	-0.019	-0.021	0.005
	(0.019)	(0.020)	(0.012)
Constant	6.226***	6.236***	5.918***
	(0.227)	(0.241)	(0.141)
Obs.	262,619	259,148	257,516
Adj. R2	0.811	0.854	0.870
FE	Seller, HS	Seller × HS	Seller × HS, Buyer

Notes: This table shows the results of estimating Equation 2. Sector s is defined according to 6-digit HS classification. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors are clustered at the seller and HS level. The higher the Bhattacharyya coefficient, the more culturally close two firms are.

Fact 3: Cultural proximity increases the likelihood of ever trading.

In this section 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 unfeasible sector combinations. This means, we drop the combinations of firms that are involved in sectors that never recorded a transaction in the data. Finally, we drop all unfeasible transactions based on distance. This is to say, we drop the combinations of firms where the seller is further away than the maximum recorded distance for the in-state buyer or vice versa.

With this sample, we estimate three different sets of specifications. The first one considers a trade dummy, defined as

$$tr\left(\nu,\omega\right)=\begin{cases}0, & n\left(\nu,\omega\right)=0\\ 1, & n\left(\nu,\omega\right)>0\end{cases},$$

whereas the second set of estimation considers the inverse hyperbolic sine (IHS) transformation and the third set of estimations considers Poisson Pseudo Maximum Likelihood (PPML) regressions. All of the estimations follow a gravity-type of estimation such that

$$y\left(\nu,\omega\right)=\iota+\iota_{\nu}+\iota_{\omega}+\delta BC\left(\nu,\omega\right)+\eta\ln dist\left(\nu,\omega\right)+\varepsilon\left(\nu,\omega\right), \tag{3}$$

where $y(\nu, \omega)$ varies across specifications. Table 3 presents the extensive margin results. Column 1 shows that the higher the Bhattacharyya coefficient, the more likely is that two given firms will trade. Columns 2 and 3 present results that combine the extensive and intensive margins through the estimation of panel regressions with the inverse hyperbolic sine transformation and PPML. They show, once again, that more cultural proximity implies a higher likelihood of trade.

Table 3: Extensive margin estimations, in-state-only sample

	(1)	(2)	(3)
Dep. Variable	Trade Dummy	Sales (IHS)	Sales (PPML)
BC	0.00088***	0.01285***	1.32505***
	(0.00008)	(0.00118)	(0.26677)
Log dist.	-0.00002	-0.00022	0.06775
	(0.00004)	(0.00057)	(0.10004)
Constant	0.00148***	0.02004***	13.30958***
	(0.00044)	(0.00623)	(1.08025)
Obs.	5,855,123	5,855,123	2,558,029
Adj. R2	0.009	0.009	.
Pseudo R2	.	.	0.539

Notes: This table shows the results of estimating Equation 3. Sample only contains in-state firms. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Regressions consider seller fixed effects and buyer fixed effects. Standard errors are clustered to the seller and buyer level. The higher the Bhattacharyya coefficient, the more culturally close two firms are. We have a smaller number of observations in Column 3 due to the dropping of observations separated by a fixed effect (Correia et al., 2019). Column 2 outcome is the inverse hyperbolic sine (IHS) of sales.

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 perform additional estimation exercises in Appendix C. We first consider an alternative measure of cultural proximity and find robust results. Then we run specifications that account for the size of the firms, where we study how differences in the sizes between small and large firms affect the relationship between cultural proximity and trade. Lastly, we see whether there are asymmetric effects due to the position of the firms in the Varna-based hierarchy.

In this section we discuss the possible mechanisms that may give rise to these findings. We focus particularly on preference-based and contracting-based mechanisms. We argue that these results cannot come 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. Yet, this channel is unlikely, as in the presence of profit maximizing firms, such firms may be competed out of the market.

Discrimination from high-caste against low-caste Varnas may again reduce trade. Yet, in Table A5 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.

We now turn to contracting-based mechanisms as an explanation. 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 in contracting frictions to find suitable trading schemes or partners. 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. 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 the following section we present a theoretical model 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.

4 Model

In this section we describe the model. First, we describe the model environment. Second, we define the equilibrium of the model. See Appendix D for further details.

4.1 Environment

Our model is based on Bernard et al. (2022), which is a quantitative firm-level production network model with heterogeneous firms and endogenous network formation. We modify their model by making firms heterogeneous in their cultural endowments as well (i.e. the probability vectors $\boldsymbol{\rho_\nu}$ over the cultural groups), which in turn we use to construct a measure of cultural proximity between firms. We allow this proximity to influence the trade cost and the matching cost.

Firms. There is a continuum of firms with technology

$$y(\omega) = \kappa_\alpha z(\omega)\, l(\omega)^\alpha\, m(\omega)^{1-\alpha}, \tag{4}$$

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

$$c(\omega) = \frac{P(\omega)^{1-\alpha}}{z(\omega)}, \tag{5}$$

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

$$p(\nu,\omega) = \mu c(\nu)\, d(\nu,\omega), \;\; d(\nu,\omega) \geq 1, \tag{6}$$

where $d(\nu,\omega)$ 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

$$n(\nu,\omega) = p(\nu,\omega)^{1-\sigma}\, P(\omega)^{\sigma-1}\, N(\omega), \tag{7}$$

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 7 we can obtain the gravity equation as

$$\ln(n(\nu,\omega)) = \iota_\nu + \iota_\omega + (1-\sigma)\log(d(\nu,\omega)), \tag{8}$$

where ι_ν and ι_ω are seller and buyer fixed effects. This gravity equation relates directly to Equation 1. Lastly, we assume the wedge is a function of different trade costs, including cultural proximity between firm CEOs due to ethnicity. Thus, we have

$$d(\nu,\omega) = \exp(\beta_1 dist(\nu,\omega) + \beta_2 BC(\nu,\omega)), \tag{9}$$

where the parameters $\beta_1 > 0$ and $\beta_2 < 0$ are trade cost semi-elasticities. The wedge will be larger the longer the geographic distance and the lower the cultural proximity. From Equation 6 we have that the higher the cultural proximity, the lower the prices, which relates to stylized fact 2. Likewise, from Equation 8 we have that the higher the cultural proximity, the higher the intermediate sales, which relates to stylized fact 1.

Households. There is a representative household that demands goods from firms. To simplify, households exhibit the same elasticity of substitution across goods σ as from firms. So, households solve

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

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

$$x(\omega) = p(\omega)^{1-\sigma} P^{\sigma-1} Y, \tag{10}$$

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

In this section 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. Then, if we work on the price index of all of the goods acquired by firm $z^{'}$, we get

$$P\left(z^{'}\right)^{1-\sigma}=\mu^{1-\sigma}\int P\left(z\right)^{(1-\alpha)(1-\sigma)}z^{\sigma-1}d\left(z,z^{'}\right)^{1-\sigma}l\left(z,z^{'}\right)dG\left(z\right), \tag{11}$$

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

$$S\left(z\right)=\frac{\left[\mu^{1-\sigma}P\left(z\right)^{(1-\alpha)(1-\sigma)}z^{\sigma-1}\right]\times}{\left[\frac{Y}{P^{1-\sigma}}D\left(z\right)^{1-\sigma}+\left(\frac{1-\alpha}{\mu}\right)\left(\int\left[d\left(z,z^{'}\right)^{1-\sigma}P\left(z^{'}\right)^{\sigma-1}S\left(z^{'}\right)\right]l\left(z,z^{'}\right)dG\left(z^{'}\right)\right)\right]}, \tag{12}$$

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

4.3 Endogenous network

In this section 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 $F\left(z,z^{'}\right)$.[9] Then, the share of seller-buyer pairs $\left(z,z^{'}\right)$ is

$$l\left(z,z^{'}\right)=\int 1\left[\ln\left(\pi\left(z,z^{'}\right)\right)-\ln\left(F\left(z,z^{'}\right)\right)-\ln\left(\epsilon\left(z,z^{'}\right)\right)>0\right]dH\left(\epsilon\left(z,z^{'}\right)\right), \tag{13}$$

[9] 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).

where $\pi\left(z,z^{'}\right)$ are the profits for seller z of selling to buyer $z^{'}$ and ϵ is an iid 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 negatively related to the cultural distance. Then

$$F\left(z,z^{'}\right)=\exp\left(\gamma BC\left(z,z^{'}\right)\right),\ \gamma<0. \tag{14}$$

From Equations 13 and 14, we see that the higher the cultural proximity, the lower the matching cost and the larger the probability of matching. This relates to stylized fact 3.

5 Estimation and calibration

Here we explain how we estimate the key parameters of the model on cultural endowments, (intensive) trade costs, and sellers' 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 $\boldsymbol{\rho_\nu}=\left[\rho_\nu(1),\ldots,\rho_\nu(452)\right]$ of belonging to each of the 452 cultural groups we observe in the data. We further assume the elements of $\boldsymbol{\rho_\nu}$ are randomly drawn from a Dirichlet distribution, such that $\rho_\nu(1),\ldots,\rho_\nu(452)\sim\mathcal{D}(\alpha_1,\ldots,\alpha_{452})$, where $\alpha_1,\ldots,\alpha_{452}>0$ are concentration parameters.[10] The probability density function for the Dirichlet distribution is

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

with the conditions

$$\rho_\nu(x)\in[0,1],\quad \sum_{x=1}^{452}\rho_\nu(x)=1,$$

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 $\boldsymbol{\alpha}=\left[\begin{array}{ccc}\alpha_1, & \ldots, & \alpha_{452}\end{array}\right]$ parameters by maximum likelihood.[11] Let $\boldsymbol{\varrho}=\{\boldsymbol{\rho_1},\ldots,\boldsymbol{\rho_N}\}$, where $\mathcal{N}$ is the total number of firms. Then, the

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

[11] For this, we use the Matlab toolboxes `fastfit` and `lightspeed` by Tom Minka. We present the estimated parameters in Figure A3 in Appendix A.

log-likelihood function is

$$\ln pr\left(\boldsymbol{\varrho}|\boldsymbol{\alpha}\right)=\mathcal{N}\ln\Gamma\left(\sum_{x=1}^{452}\alpha_{x}\right)-\mathcal{N}\sum_{x=1}^{452}\ln\Gamma\left(\alpha_{x}\right)+\mathcal{N}\sum_{x=1}^{452}\left(\alpha_{x}-1\right)\left(\frac{1}{\mathcal{N}}\sum_{\nu=1}^{\mathcal{N}}\ln\rho_{\nu}\left(x\right)\right). \quad (15)$$

Trade costs d. From Equation 9 we need an estimate for for $\{\beta_1,\beta_2\}$. We obtain estimates for these two parameters by linking the theoretical gravity equation (8) to the empirical gravity equation results (Column 1 from Table 2). Thus, we obtain $\{\beta_1,\beta_2\}=\{0,-0.02\}$.[12]

Matching cost F. From Equation 14 we need an estimate for γ. We do this in two steps. First, we go back to the extended sample from Table 3. We take the results from Column 3, that is, by PPML we estimate

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

with which we recover

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

where the hats denote estimated parameters and $\ln\widehat{n\left(z,z'\right)}$ are the predicted sales. This variable will predict what would be the sales for a pair of seller and buyer even in the case they did not actually trade in the data. In the second step we combine and rearrange equations 13 and 14, so that

$$l\left(z,z^{'}\right)=\int1\left[\ln\left(\epsilon\left(z,z^{'}\right)\right)<\ln\widehat{n\left(z,z'\right)}-\ln\left(\sigma\right)-\gamma BC\left(z,z^{'}\right)\right]dH\left(\epsilon\left(z,z^{'}\right)\right), \quad (16)$$

where we also account for the fact that $\pi\left(z,z^{'}\right)=\frac{n\left(z,z^{'}\right)}{\sigma}$ and replace $\ln n\left(z,z^{'}\right)$ by its estimated counterpart $\ln\widehat{n\left(z,z'\right)}$. We estimate this last equation with a probit regression (assuming $\epsilon\left(z,z^{'}\right)$is log-normally distributed). We find that $\gamma=-0.07$.[13]

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

[12]Even though the wedge also appears in the price equation 6 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.

[13]We present the results of the estimation in Table A1 in Appendix A.

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 $\mathcal{N} = 400$. With this we have that the total income $Y = wL = 1$.

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 $\mu_{\ln(\epsilon)}$ and (iii) the standard deviation $\sigma_{\ln(\epsilon)}$ of the link function noise distribution. We estimate these three 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. Because the link function noise distribution affects how firms match between them, then, to identify the parameters related to this distribution we must target moments that are related to the extensive margin. This being mostly a seller-oriented model, we choose to target the mean and the variance of the log-normalized number of buyers $\ln\left(\frac{\mathcal{N}_b(\nu)}{\mathcal{N}}\right)$, where $\mathcal{N}_\omega(\nu)$ is the number of buyers a seller ν has. Similarly, 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 $\ln\left(\frac{\widetilde{N}(\nu)}{\mathcal{N}_\omega(\nu)}\right)$, where $\widetilde{N}(\nu)$ is the total intermediate sales a seller ν makes.

The untargeted moments are related to the buyers. The first untargeted moment is the mean and the variance of the log-normalized number of sellers $\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)$, where $\mathcal{N}_\nu(\omega)$ is the number of sellers a buyer ω has. Then, we use 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 $\sigma_{\ln(z)} = 0.97$, $\mu_{\ln(\epsilon)} = 57.59$ and $\sigma_{\ln(\epsilon)} = 9.68$. Table 4 shows how the model-based moments fare against their empirical counterparts. When it comes to the targeted moments, the model can very closely replicate the empirical ones. For the untargeted moments, the model gets reasonably close to the data.

Table 4: Targeted and untargeted moments

Targeted moments		
	Data	Model
$mean\left[\ln\left(\mathcal{N}_b(\nu)/\mathcal{N}\right)\right]$	-9.02	-9.02
$var\left[\ln\left(\mathcal{N}_b(\nu)/\mathcal{N}\right)\right]$	1.33	1.33
$var\left[\ln\left(\widetilde{N}(\nu)/\mathcal{N}_b(\nu)\right)\right]$	3.04	3.04
Untargeted moments		
	Data	Model
$mean\left[\ln\left(\mathcal{N}_s(\omega)/\mathcal{N}\right)\right]$	-9.14	-8.52
$var\left[\ln\left(\mathcal{N}_s(\omega)/\mathcal{N}\right)\right]$	0.93	0.21
$var\left[\ln\left(N(\omega)/\mathcal{N}_s(\omega)\right)\right]$	3.12	0.59

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 $var\left[\ln\left(\widetilde{N}(\nu)/\mathcal{N}_b(\nu)\right)\right]$. 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 $var\left[\ln\left(N(\omega)/\mathcal{N}_s(\omega)\right)\right]$. The exact definition of the targeted and untargeted moments, as well as the construction of their empirical counterparts, appears in Appendix B.

6 Counterfactuals

In this section we present the results of our counterfactual exercises. We aim to quantify the welfare gains of implementing social inclusion policies through their impact on the production network. Quantifying these welfare gains is policy relevant, since they are usually not taken into account when evaluating these policies.

The model implies that welfare is measured by real wage, so welfare is $\mathcal{W} = \frac{w}{P}$. 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 activity of the economy, we measure total sales $\mathcal{S} = \sum_{\nu=1}^{N} S_\nu$, where S_ν are the total sales of firm ν. 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, $mean\left[\ln\left(\frac{\mathcal{N}_s(\omega)}{\mathcal{N}}\right)\right]$.[14]

To illustrate the magnitude of these policies, we run two extreme scenarios. First, we quantify what would happen if each firm had no chance of belonging to the same cultural group.

[14]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.

Alternatively, we can understand this as a case in which each firm belongs to its own cultural group. Thus, we go from the baseline to $BC\left(z, z^{'}\right) = 0$ 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.

Second, we quantify what would happen in every firm had exactly the same probability distribution of belonging to the cultural groups. Alternatively, we can analyze this as a case 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 takes 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.

We present the welfare gains and the average productivity gains with respect to the baseline scenario in Table 5. In the first counterfactual, when all firms are the furthest in cultural terms and the contracting frictions are the highest, average sales fall by 2.65 percent and prices increase by 1.34 percent. There are also less matches, which is reflected by an average number of buyers that falls by 0.68 percentage points and an average number of sellers that falls by 0.62 percentage points. Thus, welfare falls by 1.33 percent.

In the second counterfactual, firms become the closest in cultural terms and contracting frictions become non-existent. In this case, average sales increase by 3.29 percent and prices fall by 1.58 percent. Matching between firms also increases, reflected by an average number of buyers that grows by 0.85 percentage points and an average number of sellers that increases by 0.76 percentage points. Therefore, welfare increases by 1.61 percent.

Table 5: Counterfactual scenarios

	CF1: Each firm belongs to a different cultural group	CF2: All firms belong to the same cultural group
Welfare	-1.33%	1.61%
Average productivity	0.11%	-0.13%
Total sales	-2.65%	3.29%
Aggregate price index	1.34%	-1.58%
Average normalized number of buyers	-0.68p.p.	0.85p.p.
Average normalized number of sellers	-0.62p.p.	0.76p.p.

Notes: We present the percentage gains or losses with respect to the baseline scenario.

Besides welfare, another measure to analyze counterfactually is the average productivity. In the first counterfactual exercise, where each firm belongs to a different cultural group, average productivity slightly increases by 0.11 percent. Meanwhile, in the second counterfactual exercise, where all firms belong to the same cultural group, average productivity falls by 0.13 percent. The explanation of these results lies in whether the less productive firms are selling more or less with respect to the baseline case. We show this in Table 6. In the first counterfactual, when costs related to cultural distance increase, the less productive firms lose more sales, which lowers their weights in the aggregate and brings the average productivity up. In the second counterfactual, when cultural distance-related costs decrease, the less productive firms match more and sell more, which increases their importance in the aggregate and lowers the average productivity.

Table 6: Change in sales by productivity quartiles

	CF1: Each firm belongs to a different cultural group	CF2: All firms belong to the same cultural group
1st quartile (most productive)	-2.63%	3.26%
2nd quartile	-2.78%	3.46%
3rd quartile	-2.75%	3.46%
4th quartile (least productive)	-2.76%	3.42%

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

7 Conclusion

We aim to shed light on how cultural proximity shapes the formation of production networks and its implications for welfare. We provide evidence that proximity between firm CEOs, measured by how ethnically close are they according to India's caste and religious system, influences production networks by fostering trade, lowering equilibrium prices and increase the probability of matching between sellers and buyers. We then develop a production network model that we estimate using the data from our empirical section. Lastly, we use the model to analyze policy counterfactuals that foster social inclusion in India.

Our paper is the first attempt to closely study how social relationships influence firm-level decisions and quantify its importance for welfare, in both empirical and quantitative aspects. In terms of policy, our model allows us to properly evaluate social inclusion policies that take into account its influence on production networks. This dimension of these policies is usually largely ignored and we aim to show that this aspect is key for their implementation.

References

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.

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

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.

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.

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.

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.

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.

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

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. (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.

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

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

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.

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.

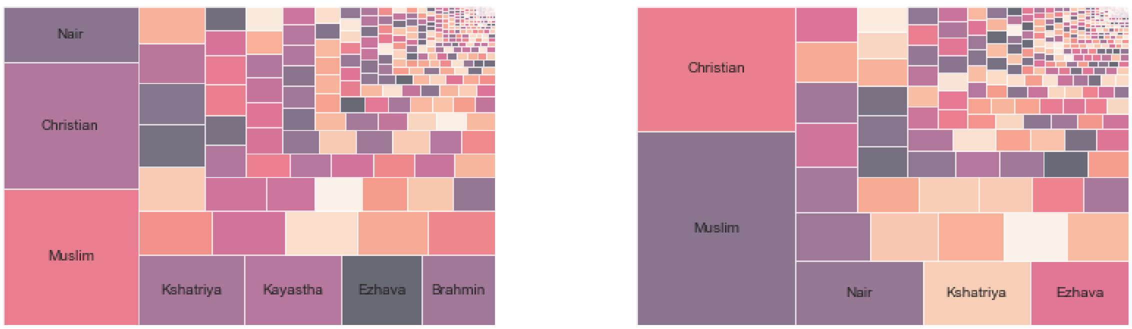
Taschereau-Dumouchel, M. (2019). Cascades and fluctuations in an economy with an endogenous production network.

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

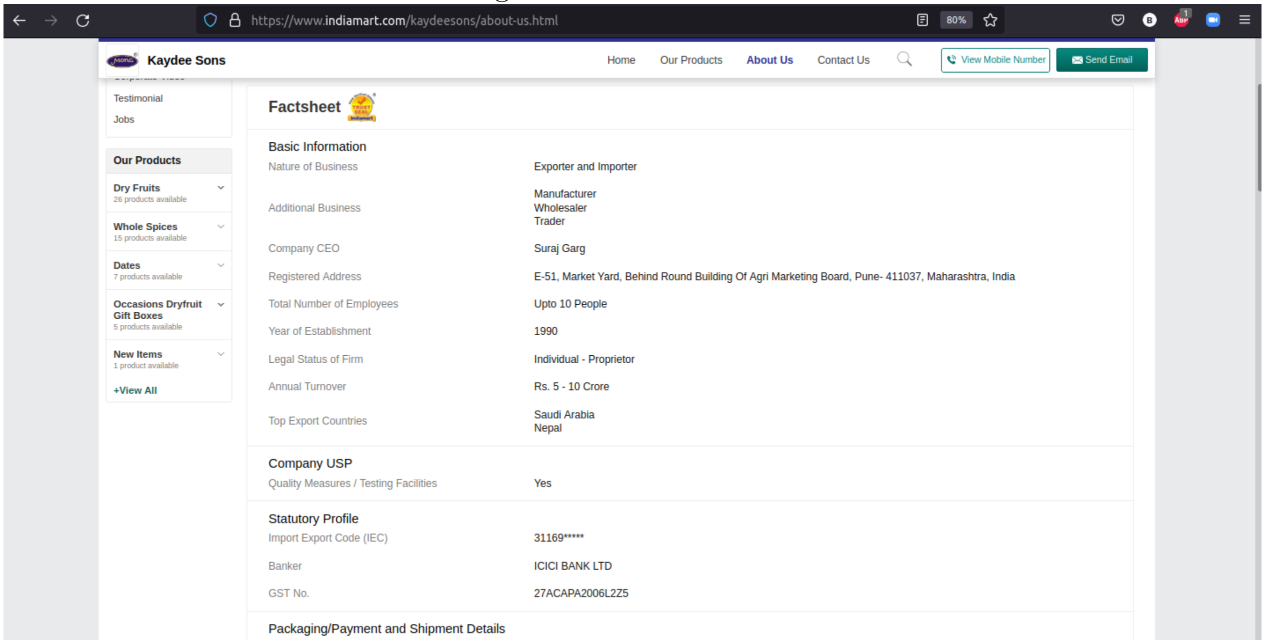
Figure A1: Probability-weighted sales and purchases across cultural groups

(a) Sales (b) Purchases



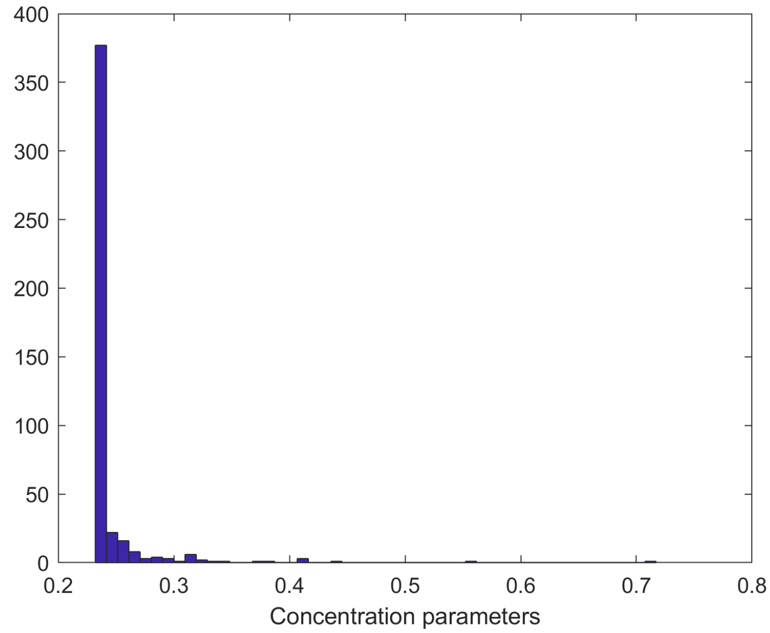
Notes: Figure shows the decomposition of the probability-weighted sales and purchases across the 452 cultural groups in our dataset.

Figure A2: IndiaMART



Notes: This vignette shows an example on how a firm in the website IndiaMART looks like. This is the factsheet firm called *Kaydee Sons* that sells dry fruits, dates, spices, seeds, walnut kernels, gift pack, and fresh apricot. We scrape all the variables included in this vignette, where the name of the company CEO and the GST number are the most important ones.

Figure A3: Histogram of estimated concentration parameters for Dirichlet distribution



Notes: Estimated concentration parameters for a Dirichlet distribution according to the maximum likelihood estimation from equation 15.

Table A1: Second stage estimation for matching cost

	(1)
	Probit
BC	0.076***
	(0.010)
$\ln\widehat{n\left(z,z'\right)}$	0.095***
	(0.001)
Constant	-3.343***
	(0.010)
Obs.	5,855,126
Pseudo R2	0.053

Notes: This table shows the results of estimating equation 16. We winsorize $\ln\widehat{n\left(z,z'\right)}$ at 1% and 99%. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors in parentheses.

B Targeted and untargeted moments

B.1 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.

B.2 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.

C Additional regressions

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 seller and buyer have exactly equal probability distributions, while it will be higher the more different the two probability distributions are.[15] 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 $\varepsilon \to 0^{+}$ such that KL_{sym} is well-defined.

Tables A2, 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 the main text.

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

Table A2: Estimations, intensive margin, Kullback-Leibler

	(1)	(2)
Dep. Variable	Log Sales	Log Transactions
KL_{sym}	-0.004**	-0.001*
	(0.002)	(0.001)
Log dist.	-0.001	-0.033***
	(0.017)	(0.007)
Constant	13.346***	1.223***
	(0.193)	(0.085)
Obs.	32,773	32,773
Adj. R2	0.439	0.309

Notes: This table shows the results of estimating Equation 1. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Regressions consider seller fixed effects and buyer fixed effects. Standard errors are clustered to the seller and buyer level. A higher Kullback-Leibler divergence means two firms are socially farther away.

Table A3: Estimations, prices, Kullback-Leibler

	(1)	(2)	(3)
Dep. Variable	Log Prices	Log Prices	Log Prices
KL_{sym}	0.005***	0.004***	0.002*
	(0.001)	(0.001)	(0.001)
Log dist.	-0.019	-0.021	0.005
	(0.019)	(0.020)	(0.012)
Constant	6.137***	6.163***	5.873***
	(0.218)	(0.231)	(0.145)
Obs.	262,619	259,148	257,516
Adj. R2	0.812	0.854	0.870
FE	Seller, HS	Seller × HS	Seller × HS, Buyer

Notes: This table shows the results of estimating Equation 2. Sector s is defined according to 6-digit HS classification. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Standard errors are clustered at the seller and HS level. A higher Kullback-Leibler divergence means two firms are socially farther away.

Table A4: Extensive margin estimations, in-state-only sample, Kullback-Leibler

	(1)	(2)	(3)
Dep. Variable	Trade Dummy	Sales (Inverse hyperbolic sine)	Sales (PPML)
KL_{sym}	-0.00004***	-0.00055***	-0.05721***
	(0.00000)	(0.00005)	(0.01353)
Log dist.	-0.00002	-0.00023	0.06671
	(0.00004)	(0.00057)	(0.10022)
Constant	0.00224***	0.03123***	14.45027***
	(0.00045)	(0.00631)	(1.17130)
Obs.	5,855,123	5,855,123	2,558,029
Adj. R2	0.009	0.009	.
Pseudo R2	.	.	0.537

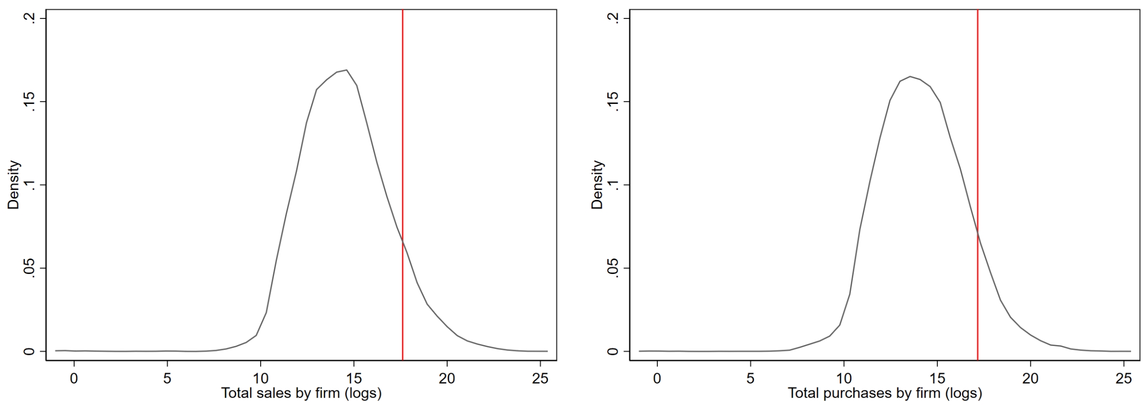
Notes: This table shows the results of estimating Equation 3. Sample only contains in-state firms. ***, ** and * indicate statistical significance at the 99, 95 and 90 percent level respectively. Regressions consider seller fixed effects and buyer fixed effects. Standard errors are clustered to the seller and buyer level. A higher Kullback-Leibler divergence means two firms are socially farther away. We have a smaller number of observations in Column 3 due to the dropping of observations separated by a fixed effect (Correia et al., 2019).

C.2 Size of firms

In this section we analyze how the size of the firms interacts with the cultural proximity. We classify a firm as large if it belongs to the top 10 percentile of sales or the top 10 percentile of purchases. We show these distributions in Figure A4.

Figure A4: Distribution of sales and purchases across firms

(a) Sales (b) Purchases



Notes: Vertical lines denote the 90th percentile cutoff. Firms to the right of the threshold are considered large.

Table A5 presents the results to a modified version of the intensive margin regression from Equation 1. We consider the interaction of the Bhattacharyya coefficient with firm size indicators. We find that cultural proximity matters when the seller is of the same size of the buyer (small firm sells to small firm or large firm sells to large firm). Moreover, the cultural proximity effect is stronger when the two parts of the transaction are large firms.

We also find that cultural proximity is less relevant when firms are of different size. An example of this would be the anonymized transaction in which a small buyer incurs. If a small buyer needs to buy supplies from one of the biggest companies in India, it is unlikely the buyer will care about which cultural group the CEO of the large seller is from.

Table A5: Estimations, intensive margin, size of firms

	(1)	(2)
Dep. Variable	Log Sales	Log Transactions
$BC \times \mathbb{I}_{\nu_S \omega_S}$	0.199***	0.063**
	(0.058)	(0.029)
$BC \times \mathbb{I}_{\nu_L \omega_L}$	0.280***	0.082**
	(0.067)	(0.032)
$BC \times \mathbb{I}_{\nu_L \omega_S}$	-0.029	-0.013
	(0.055)	(0.029)
$BC \times \mathbb{I}_{\nu_S \omega_L}$	-0.123	0.003
	(0.076)	(0.039)
Obs.	32,773	32,773
Adj. R2	0.440	0.310

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. Regressions consider seller fixed effects and buyer fixed effects. Standard errors are clustered to the seller and buyer level. The higher the Bhattacharyya coefficient, the more culturally close two firms are. The subindex that accompanies ν denotes the size of the seller, while the subindex that accompanies ω denotes the size of the buyer. S denotes a small firm and L denotes a large firm.

C.3 Asymmetric effects

In this part we study whether there are asymmetric effects in transactions in which one firm is placed higher than the other according to the Varna-based hierarchy. For this we generate indicators based on which is the Varna or religion for which a firm has the highest probability of belonging to. 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 estimations.

Table A5 shows the results of a modified version of the intensive margin regression from Equation 1. We consider interactions of the Bhattacharyya coefficient with hierarchy indicators. The baseline category is that both firms belong to the same hierarchy. In first place, we find the baseline coefficient is very similar to those of Table 1. In second place, we find there is no additional effect of cultural proximity when firms are placed differently in the hierarchy. Thus, we conclude that there are no asymmetric effects caused by vertical discrimination across cultural groups.

Table A6: Estimations, intensive margin

	(1)	(2)
Dep. Variable	Log Sales	Log Transactions
BC	0.098***	0.032*
	(0.037)	(0.019)
$BC \times \mathbb{I}_{\nu_H \omega_L}$	0.011	0.070
	(0.130)	(0.061)
$BC \times \mathbb{I}_{\nu_L \omega_H}$	0.003	-0.073
	(0.142)	(0.068)
Obs.	31,305	31,305
Adj. R2	0.442	0.309

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. Regressions consider seller fixed effects and buyer fixed effects. Standard errors are clustered to the seller and buyer level. The higher the Bhattacharyya coefficient, the more culturally close two firms are. 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.

D Derivations

In this section we include details about the derivations of the 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

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

Merge the first and third constraints, such that

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

Rewrite the minimization problem, such that

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

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

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

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

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

We divide both first order conditions, such that

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

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

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

which is the demand of firm ω from variety ν, where $n(\nu,\omega) = p(\nu,\omega)\, m(\nu,\omega)$ is the expenditure of ω on variety ν, and $N(\omega) = P(\omega)\, m(\omega)$ is the total expenditure of firm ω. Also, the expression for unit cost of production is

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

where labor is the numeraire good, so $w = 1$. 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, such that

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

The first order condition is

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

where $\mu = \frac{\sigma}{\sigma-1}$ is the markup.

D.2 Households

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

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

The first order condition with respect to firm ω is

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

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

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

We then divide both first order conditions, such that

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

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

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

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

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

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

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

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

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

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.