Identity, Market Access, and Demand-led Diversification*

Sampreet S. Goraya† Akhil Ilango‡

[*Preliminary and Incomplete: Please do not circulate*]

Abstract

We present a model and reduced-form evidence to examine and quantify *demand-led diversification* in firms, wherein firms diversify their workforce to access diverse consumer markets. Consumer preferences are influenced by both product quality and producer identity. Firms strategically exploit these preferences by hiring workers from the targeted consumer group, expanding their customer base and revenues. The decision to hire outside their group depends on market demand from other groups and associated hiring costs. Diversified firms with a varied workforce can tap into multiple markets, achieving scale. We apply this framework to caste norms in rural India, using variations in local rainfall intensity as an exogenous demand shifter. Rainfall increases spending among low-ranked caste (LC) consumers, fostering growth in local LC-owned firms. Firms from other castes adjust their workforce composition towards LC employees. Matching the estimated revenue elasticity, we find a substantial preference for identity in demand, resulting in smaller firm size and homophily in hiring practices. A lower cost of cross-caste hiring incentivizes firms to cater to ethnically distant markets, enhancing overall consumer welfare.

Keywords: Identity, Market access, Demand, Firm size, Diversity, Trade, The caste system.

*This is a substantially revised version of the paper titled Social Norms, trade and firms size.

†Princeton University & Stockholm School of Economics. Email: sg5077@princeton.edu

‡Indian Institute of Management Bangalore. Email: akhil.ilango@iimb.ac.in

1 Introduction

Firm growth is key for economic development. Existing literature has attributed cross-country differences in firm size to credit constraints, labour market regulations, taxes and subsidies, poor infrastructure and limits to delegations, among others. This paper focuses on the impact of consumer identity–defined as consumers' sense of belonging to a group–on market access, hiring practices, and firm size. Identity plays a pivotal role in economic decisions, influencing social interactions and consumption patterns. As consumer preferences shift based on identity, firms must align hiring strategies to appeal to a diverse consumer base. However, the interplay between consumer preferences for in-group products and producers and a firm's hiring strategy remains unexplored. Do these interactions influence a firm's entry decision, its access to diverse markets and consequently its size? And do they have macroeconomic implications by affecting industry competition and consumer welfare? Is there a role for policy? In this paper, we answer these questions both theoretically and quantitatively.

We document and quantify demand-led diversification: a phenomenon of choosing diverse workforce composition to gain access to diverse consumer markets. We posit that adherence to group identity is reflected in consumers favouring products from firms owned by individuals within the same group, be it based on race, gender, ethnicity, or caste. Consequently, firms proactively invest in shaping consumer demand by recruiting employees from the target consumer group. As a result, the boundary of a firm is determined by the degree of cross-group trade, which in turn depends on the cost of hiring employees from the target consumers' group.

To formalize this argument, we build a heterogeneous firm general equilibrium model that explains the decision of firms to sell to diverse groups and the composition of employees for small and large firms. Each group has a unique and distinguishable identity, characterised by their social distance from each other, and is heterogeneous in their economic size – some are rich and some are poor. Firms are heterogeneous in the quality of the products they produce and the identity of the owner. Both of these attributes determine a product's demand. Conditional on the quality, consumers have a taste for the goods produced by socially closer groups relative to distant groups. While accessing consumers of different groups, we assume that firms are required to hire a fixed number of employees from the target consumer group. We find that, in equilibrium, firms that sell to many diverse groups are positively selected, tend to be larger, and have a diverse workforce.

Firms that indulge in cross-group trade can also invest in increasing taste for their product by hiring more employees of the target group. However, becoming socially closer to one group may come at the cost of becoming socially farther away from another. Therefore, the firm's decision to enter a market not only depends on a product's demand from the target group but also the distribution of product demands of all relevant groups. We formally show that the optimal ethnic location for the firm is a weighted average of the ethnic distance between groups, where weights are proportional to product demand and the cost of moving closer to each market. For instance, if a certain group has a large product demand, then firms would prefer to position themselves closer to this group. However, if the cost of moving ethnically closer to this group is large, then the firm would hesitate to hire and move closer. Large firms from small groups have the incentive to be closer to richer groups. Taken together, the model provides one rationalization behind why firms would actively pursue workforce diversification even in the absence of any supply-side benefits or in the presence of any upside risk such as workplace conflicts. A direct implication of the model is differences in consumers' concern for identity can trigger cross-regional differences in firm size and homophily in hiring.

Next, we test the model's predictions in the context of caste norms in rural India. Caste is inherited at birth and determines one's social identity. Historically, the caste system has been restricting inter-caste interactions and promoting discriminatory practices towards low-ranked communities (LC, hereafter) in the hierarchical system. Despite immense socioeconomic changes in the last few decades, caste remains a salient feature of Indian society (Munshi, 2019).[1]

We use the 2006-07 Micro, Small, and Medium Enterprises (MSME) dataset, incorporating caste information for employers and employees, as well as product revenues and quantities – features lacking in other widely used firm-level datasets (Goraya, 2023). Augmented by household-level surveys, our findings align with prior literature, indicating lower incomes and consumption among LC consumers and smaller firms owned by LC compared to HC counterparts. Our analysis further reveals two novel patterns: first, homophily in employee hiring, with around 75% belonging to the same caste as the employer, and second, larger firms having a more diversified workforce. To establish causation between demand and the caste composition of the workforce, we use local rainfall variation as a proxy for exogenous shifts in demand, especially among LC households.

To establish rainfall as an asymmetric demand shifter, we examine its impact on households' Monthly Per-Capita Expenditure ($mpce$), employing district $\times$ year, and caste fixed effects.[2] A 1% increase in rainfall corresponds to a 10% average rise in $mpce$ for LC consumers relative to the HC consumers. This consumption increase is consistent across non-agricultural sectors, including manufacturing, services, and durables. Although higher rainfall primarily affects input supply and production in the agricultural sector, it indirectly stimulates demand in other sectors, indicating a notable indirect effect on firms' output in

[1] Several papers have documented channels through which caste affects economic outcomes such as allocation of capital (Banerjee and Munshi, 2004; Fisman, Paravisini and Vig, 2017; Goraya, 2023), migration and network insurance (Munshi and Rosenzweig, 2016), labour markets (Madheswaran and Attewell, 2007; Oh, 2023; Cassan, Keniston and Kleineberg, 2021), health and education outcomes (Munshi and Rosenzweig, 2006; Spears and Thorat, 2019); and female seclusion norms (Agte and Bernhardt, 2023).

[2] See also Jayachandran (2006); Adhvaryu, Chari and Sharma (2013); Chaurey (2015); Kaur (2019); Santangelo (2019); and Gupta (2020) for notable work on the effects of rainfall on the Indian labour market.

diverse sectors due to heightened demand. To test the existence of caste linkages in LC firms' demand, we next estimate the effect of higher rainfall on firm-level outcomes, employing district $\times$ year, sector or product $\times$ year, and caste fixed effects. We find that corresponding to the observations about *mpce*, there is an increase in the annual revenue of LC firms by 17.2%, relative to HC firms, for a 1% increase in rainfall. We find that the firms increase their use of intermediate inputs as well.

In assessing workforce adjustments to demand changes, we employ district, sector, and caste fixed effects due to caste composition availability for only one year. We offer three lines of evidence: firstly, among low-caste (LC) and medium-caste (MC) firms, high rainfall regions exhibit a low share of high-caste (HC) employees and a high share of LC and MC employees. Secondly, in competitive markets, HC firms tend to hire more LC employees when LC consumer demand or rainfall is high compared to HC-dominated markets.[3] Thirdly, examining sectoral differences in customer-facing industries, like Repair and maintenance and Services, reveals that HC-owned firms hire more LC employees in contact-intensive sectors with high LC demand. This cumulative evidence suggests that changes in demand composition are driving shifts in employee composition.

We provide evidence to rule out alternative mechanisms that may lead to the same economic outcomes. Using the information on village-level population shares, we show that potential geographic segregation by castes is not driving the results. Further, using data on input and output prices, we show that changes in product quality are unlikely to be driving the demand effects. Further, the insignificant effect of rainfall on product prices within our sample also shows that the rise in revenue is driven by a rise in quantity. In addition, to address concerns on the supply-side caste linkages (e.g., through loans from caste networks), we test how the amount of informal (or formal) loans varies by low and high rainfall regions and find no systematic evidence. Taken together, these observations indicate that the taste for identity in demand is the primary driver of our results.

Finally, we engage in a quantitative analysis of our model, undertaking simulations to explore the macroeconomic implications of various policies. This necessitates the precise determination of four parameters: the taste for identity, fixed operational costs, fixed trading costs outside one's caste, and the scale parameter of productivity distribution. These parameters are jointly estimated, with specific moments associated with each parameter aiding in their identification. To gauge the taste for identity parameter, we align the partial equilibrium response of the economy with a demand shock that exclusively impacts LC and MC groups. Specifically, we calibrate the taste for identity parameter by matching the revenue elasticity of LC and MC-owned firms, relative to HC firms, within our framework.[4]

[3] HC-dominated markets are defined by sector-district pairs where the share of firms/revenues/employees for HC-owned firms is above a certain cut-off. We show that results are consistent across various cut-off definitions. These results also highlight that changes in the composition of HC firms' workforce are not primarily driven by fluctuations in LC labor supply induced by rainfall. Despite potential changes at the district level, there is no clear rationale for expecting them to align with variations in competition within specific industries.

[4] We demonstrate a monotonic increase in the relative revenue elasticity of LC and MC-owned firms with

We leverage the revenue share of the bottom 50% and top 10% of the firm size distribution to calibrate the fixed cost of production and the scale parameter of the productivity distribution. Additionally, we employ the share of firms engaging in cross-caste hiring to determine the fixed cost of trading as we do not have data on cross-caste trade. Notably, in our model, the necessity for firms to hire employees from the same caste they are selling to implies an equivalence between the share of firms hiring and trading outside one's caste. While this assumption may not precisely mirror data, we find empirical support, with the share of firms hiring outside one's caste approximating 38% in our data, closely mirroring the probability of trading outside one's caste documented in recent papers.[5]

The calibrated version of the model finds a substantial taste for identity in the economy with demand decaying at a rate of 30.8% over social distance between castes. Employing this calibrated model, we show that taste for identity makes firms' markets socially local. Firms are less likely to trade with socially distant castes and thus have lower incentive to hire employees from other castes. This means that aggregate homophilly in the economy is high and the average firm size is small.

In our policy experiment, we investigate the impact of a policy entailing subsidies for cross-caste hiring on consumer welfare. Our findings indicate that this subsidy effectively diminishes firms' costs associated with cross-caste trade, acting as a de facto reduction in the cost of market access. Consequently, firms are incentivized to engage in selling to socially distant castes. This policy intervention yields two primary consequences: firstly, an overall reduction in homophily across the economy, as an increased number of firms partake in cross-caste trade; secondly, the entry of firms into diverse markets augments the variety of goods available to consumers, increasing overall welfare. These policy counterfactual contributes to an imminent question in front of the organizations and policymakers around the world where diversity matters; should we or should we not encourage diversity in hiring practices?[6] We show that demand-side considerations are an important piece to this riddle.

Literature Review. First, this paper is related to the literature on ethnic networks and trade (Rauch, 2001; Rauch and Trindade, 2002; Combes, Lafourcade and Mayer, 2005; Felbermayr and Toubal, 2010; Aker, Klein, O'Connell and Yang, 2014; and Desmet, Gomes and Rico, 2023).[7,8] Anderson (2011) finds that rural groundwater markets are segmented by caste and

consumers' taste for identity, providing a robust foundation for our parameter estimation.

[5]Using a detailed firm-firm sales data, Boken, Gadenne, Nandi and Santamaria (2022) estimates that the probability of trading within caste is twice the probability of trading outside one's caste. We also provide results where we exactly match this probability of trading to calibrate the cost of trading.

[6]See, recent report published by OECD (2020). Many European countries have some policies in place to push diversity in hiring in the private sector. Moreover, the idea that a diverse workforce allows firms to expand their customer base has echoed in Hunt, Prince, Dixon-Fyle and Yee (2018).

[7]A related literature pioneered by McCallum (1995) finds substantial home bias in trade.

[8]A separate body of research has documented that ethnic fragmentation can impede the provision of public goods (Alesina, Baqir and Easterly, 1999; Alesina and La Ferrara, 2005), diminish social capital (Alesina and La Ferrara, 2000), elevate conflicts and impede growth (Montalvo and Reynal-Querol, 2005a,b). Furthermore, economic outcomes worsen when there is greater overlap between ethnicity and culture (Desmet, Ortuño-

Emerick (2018) shows caste as an impediment to the exchange of seeds. Nagavarapu and Sekhri (2016) finds that Scheduled Castes (LC) are more likely to buy grains when facing LC delivery agents. Fujiy, Khanna and Toma (2022) and Boken, Gadenne, Nandi and Santamaria (2022) highlight that firms from the same caste are more likely to trade. This paper highlights the caste linkages in firms' output market, that is with their consumers, and provides causal evidence of an asymmetric transmission of demand shocks across firms. Further, we provide both theoretical and empirical results on how demand-side tastes may spillover to firms' input markets, specifically, the labor market.[9]

Second, this paper adds to the growing body of evidence on the importance of demand-side factors for firm growth (Foster, Haltiwanger and Syverson, 2008, 2016; Hottman, Redding and Weinstein, 2016; Einav, Klenow, Levin and Murciano-Goroff, 2021; and Bernard, Dhyne, Magerman, Manova and Moxnes, 2022). Several papers have highlighted the role of market access in firm entry and growth (Allen, 2014; Atkin, Khandelwal and Osman, 2017; Jensen and Miller, 2018; Hjort and Poulsen, 2019) in the context of emerging markets. Further, gender and race are linked to firms' demand (see Hardy and Kagy, 2020 for female-owned firms and Tan and Zeida, 2023 for black-owned firms). Our paper provides evidence on the demand-side ethnic linkages that restrict firms' market access and reduce firm size. Further, we show that firms use workforce diversity as a tool for market access, expanding the body of evidence on why firms tend to be small and grow slowly in developing countries.

Third, this paper contributes to the literature that links the ethnic/religious composition of the labor force and firm scale and profitability. Hjort (2014); Afridi, Dhillon and Sharma (2020), and Ghosh (2022) show that ethnic heterogeneity lowers firms' productivity, whereas Bhagavatula, Bhalla, Goel and Vissa (2018) and Brinatti and Morales (2021) find that better performance and worker diversity are positively linked. In our model, we show that more ethnically diverse firms are bigger and more profitable because they can appeal to diverse consumer markets, thus highlighting the demand channel.[10]

Fourth, this paper is connected to the literature that studies trade across interdependent markets (see, Albornoz, Pardo, Corcos and Ornelas, 2012; Schmeiser, 2012; Chaney, 2014; Morales, Sheu and Zahler, 2019; and Alfaro-Urena, Castro-Vincenzi, Fanelli and Morales, 2023). Relative to these papers, our contribution is to provide a theory that selling to a certain market may affect a product's appeal in other relevant markets. Further, we show how the interdependence between different markets shapes employee composition within

Ortín and Wacziarg, 2012).

[9]Beyond the caste system, Alsan, Garrick and Graziani (2019) document the preferences for the same race doctors by patients in the US healthcare sector and Takeshita, Wang, Loren, Mitra, Shults, Shin and Sawinski (2020) shows that patients report better experience when matched with same race doctors.

[10]Hiring from the network may have supply-side benefits. A large literature has documented advantages and disadvantages of in-group hiring (see, Montgomery, 1991; Munshi, 2003; Bertrand and Schoar, 2006; Karlan, Mobius, Rosenblat and Szeidl, 2009; Beaman and Magruder, 2012; Beaman, 2016; Chandrasekhar, Morten and Peter, 2020; Heath, 2018; Dhillon, Iversen and Torsvik, 2021; Carranza, Garlick, Orkin and Rankin, 2022 and Fernando, Singh and Tourek, 2023).

a firm.

The remainder of this paper is organized as follows: Section 2 presents a framework of spatially and ethnically heterogeneous firms, and derives testable equilibrium implications. Section 3 provides a brief description of the Caste system in India. Section 4 introduces the data. Section 5 discusses our empirical strategy and documents the empirical evidence. Section 6 outlines the quantitative analyses. Section 7 concludes by discussing some avenues for future research.

2 Theoretical Framework

In this section, we describe a model of within-district sales, in a district populated by members of different castes. There are $\mathcal{S}$ castes in the district and the ethnic distance between two castes denoted by s and s' is defined as Euclidean distance $d_{ss'}$ and by symmetry $d_{ss'} = d_{s's}$. We assume that $0 \leqslant d(s,s^{'}) \leqslant 1$, where $d_{ss'} = 0$ means zero ethnic distance and $d_{ss'} = 1$ means s and $s^{'}$ are ethnically most distant castes. We will define the ethnic distance more precisely in Section 2.3. We now describe the household sector and the production sector.

2.1 Households

For each caste, there is a representative household denoted by s. It has a total labour endowment of L_s units.[11] The household has non-homothetic preferences over two types of goods: homogeneous goods and differentiated varieties. In particular, the household solves the following problem

$$\mathcal{U}(C_{H,s}, C_{D,s}) = \max_{C_{H,s}, \{c(z(\omega),s,s^{'})\}_{j\in\Omega_s}} a\log(C_{H,s} - \overline{H}) + (1-a)\log C_{D,s} \tag{1}$$

$$C_{D,s} = \left[\sum_{\Omega_s} q(z(\omega),s,s^{'})c(z(\omega),s,s^{'})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{2}$$

$$I_s = w_e L_s + \Pi \tag{3}$$

$$I_s \geq P_H C_{H,s} + \sum_{\Omega_s} p(z(\omega),s,s^{'})c(z(\omega),s,s^{'}), \tag{4}$$

where $C_{H,s}$ is the demand for homogeneous goods and $C_{D,s}$ is the demand for high-quality differentiated goods. $C_{D,s}$ is a bundle of differentiated varieties produced by firms of different castes at a cultural distance $d_{ss^{'}} \in [0,1]$. Further, we assume that households have CES preferences over endogenous Ω_s differentiated varieties, where σ denotes the elasticity of substitution between varieties. Each variety ω has two attributes, quality $z(\omega)$ and the caste of producer $s^{'}$. The $c(z(\omega),s,s^{'})$ denotes the units of consumption by household s of good quality $z(\omega)$ and produced by caste $s^{'}$, $q(z(\omega),s,s^{'})$ denotes the utility derived per unit

[11] The income differences across castes are driven by L_s and this captures both quantity and quality of labor.

good consumed, $p(z(\omega), s, s^{'})$ is the price. Further, it is assumed that a household needs a minimum amount of homogeneous goods $\overline{H}$ to survive. This implies that as households get richer their share of expenditure on homogeneous goods declines and expenditure on differentiated varieties increases. Household total income is denoted by I_s which is composed of total wage income and net profits from firms that belong to the same caste. The optimal expenditure and the demand for a variety with quality $z(\omega)$ is given by an iso-elastic demand curve:

$$P_H C_H = P_H \overline{H} + a(I_s - P_H \overline{H}) \tag{5}$$

$$\sum_{\Omega_s} p(z(\omega), s, s^{'}) c(z(\omega), s, s^{'}) = (1-a)(I_s - P_H \overline{H}) \tag{6}$$

$$c(z(\omega), s, s^{'}) \ = y(z(\omega), s, s^{'}) \ = \ q(z(\omega), s, s^{'})^{\sigma} p(z(\omega), s, s^{'})^{-\sigma} \kappa_s \tag{7}$$

The representative household first allocates $P_H \overline{H}$ amounts of income to $\overline{H}$ units of homogeneous good and then allocates the remaining income to the differentiated goods proportional to their weights in the utility function. Any variety with a taste shifter $q(z(\omega), d_{ss'})$ differs by their quality $z(\omega)$ and the distance $d_{ss'}$ – a measure of cultural proximity or farness – to the caste by which goods are produced. $\kappa_s = C_{D,s} P_{D,s}^{\sigma} = (1-a)(I_s - \bar{H}) P_{D,s}^{\sigma-1}$ denotes the caste-specific aggregates. Here, we can see that the demand for differentiated goods depends on the household's income. For instance, low castes, which are also poorer will demand lower quantities of differentiated varieties. We assume $q(z(\omega), s)$

$$q(z(\omega), s, s^{'}) = z(\omega) \Psi(d_{ss'}),$$

Further, $\Psi(d_{ss'})$ captures the dislike for the goods produced by castes that are farther away from my caste. The taste shifter satisfies the following properties: $\frac{\partial q}{\partial z}$ ¿ 0, the taste is higher for high-quality goods, and $\frac{\partial q}{\partial d_{ss'}}$ ¡ 0, the taste is lower for goods produced by castes that are farther away from consumers' caste.

2.2 Production Sector

2.2.1 Homogeneous good sector

There is a representative firm from each caste s in the homogeneous good sector and this good is used as the numeraire. It is produced under constant returns to scale with one unit of labor producing 1 unit of homogeneous good. Its price is set equal to one so that if caste s produces this good, the wage is normalized to one. We assume that a is large enough such that all castes produce the homogeneous good and all wages are equal across castes and equal to one.

2.2.2 Differentiated Good Sector

We assume that differentiated products are produced by a continuum of firms. Therefore, we use z as a marker of product quality. In this sector, firms produce with a Cobb-Douglas production technology with constant returns to scale, $y(z, s) = \mathcal{F}_s(\{\ell_e(s')\}_{s' \in S})$. The labour is the only input in production. The firms consider each ethnic market, say of caste s' at a distance $d_{ss'}$, as a separate market. In order to access this market, firms need to pay a fixed cost of $f^d_{ss'}$ every period.[12] Given consumers' demand, a firm with product quality z, and caste s solves the following profit maximization in each ethnic market at a cultural distance $d_{ss'}$.

$$
\begin{aligned}
\pi(z, s', s) = & \max_{p(z,s',s), \ell(z,s)} p(y(z, s', s))y(z, s', s) - \mathcal{C}(s)y(z, s', s) - f^d_{ss'} \\
s.t. \quad & y(z, s', s) = q(z, s', s)^{\sigma} p(z, s', s)^{-\sigma} \kappa_{s'} \\
& y(z, s', s) = \mathcal{F}_s(\{\ell_e(s'')\}_{s'' \in S})
\end{aligned}
$$

The first order condition with respect to $p(z, s', s)$, gives us the standard results

$$p(z, s', s) = \frac{\sigma}{\sigma - 1} \mathcal{C}_s. \tag{8}$$

The firm charges the same price to all caste groups and it is a product of firm-specific marginal cost $\mathcal{C}_s$ and a markup $\frac{\sigma}{\sigma-1}$. As usual, markup is decreasing in the elasticity of substitution. Further, firms sell more to culturally proximate groups, such that

$$y(z, s', s) = q(z, s', s)^{\sigma} \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_s \right)^{-\sigma} \kappa_{s'}, \quad r(z, s', s) = q(z, s', s)^{\sigma} \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_s \right)^{1-\sigma} \kappa_{s'},$$
$$\pi(z, s', s) = q(z, s', s)^{\sigma} \frac{1}{\sigma} \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_s \right)^{1-\sigma} \kappa_{s'} - f^d_{s's}$$

As there are fixed costs, there is a threshold quality $z^*_{s's}$ above which firm owners of caste s sell to caste s'.

2.2.3 Firm Scale

At the extensive margin, firms only sell to a caste if the profits are higher than the fixed cost of selling to that caste. At the intensive margin, the exact proportion of the sales to each caste depends on two opposite forces. First, as firms sell culturally farther away, their products become less tasteful, thus the sales to culturally distant castes are lower. However, if the size of the culturally distant groups is increasing then the product demand goes up. The size of

[12] We allow for fixed cost to be caste-dependent.

the firm from caste s producing variety ω is

$$R(z,s) = \sum_{s} r(z,s',s) = \left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma} \sum_{s} \Psi(d_{ss'})^{\sigma}\kappa_{s'}, \tag{9}$$

$$\Pi_D(z,s) = \sum_{s} \pi(z,s',s) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma} \sum_{s} \Psi(d_{ss'})^{\sigma}\kappa_{s'} - \mathcal{F}^d_s. \tag{10}$$

2.3 Endogeneous Ethnic Proximity

We allow firms to position themselves closer or farther away from certain castes. This feature allows entrepreneurs to overcome the disadvantage of being born into a certain caste group. However, the incentive to be closer to certain castes is to take advantage of their large size but it comes at the cost of losing the demand from their own group. Also, there are costs associated with changes in firms' identity. These costs are paid in wages (we assume that firms can hire sales representatives from other castes). We assume that all castes are placed in $\mathcal{R}^{\mathcal{N}}_{+}$ Euclidean space, where $\mathcal{N} = \mathcal{S}$, and $\mathcal{S}$ is the number of castes in the region.

Let us define $\mathcal{X}_s = \{\mathcal{X}_{s,1}, ..., \mathcal{X}_{s,\mathcal{N}}\}$, a column vector, contains the Cartesian coordinates of caste s in the $\mathcal{N}$ dimensional space.[13] The euclidean distance between two castes represented by $\mathcal{X}_s$ and $\mathcal{X}'_s$ is $d^2_{ss'} = \sum_{k=1}^{\mathcal{N}}(d_{ss',k})^2$, where $d_{ss',k} = \mathcal{X}_{s,k} - \mathcal{X}_{s',k}$ is the L1 distance in the k^{th} dimension. Further, let us define the distance moved by a firm owner that was born in a caste group s by a $\mathcal{N} \times 1$ column vector $\Delta\mathcal{X}_s = \{\Delta\mathcal{X}_{s,1}, ..., \Delta\mathcal{X}_{s,\mathcal{N}}\}$, where $\Delta d^2_s = \sum_{k=1}^{\mathcal{N}}(\Delta\mathcal{X}_{s,k})^2$.[14] We assume firms pay a caste-dependent moving cost that is given by a function $\Phi(\Delta\mathcal{X}_s;\Gamma_s)$, where Γ_s is $\mathcal{N} \times 1$ column vector of parameters that disciplines the costs of moving. Here, we allow for the possibility that LC firms may pay higher or lower to move closer to HC consumers than HC firms pay to move closer to LC consumers.[15] Profits of the firm owner s that chooses to move is given by

$$\Pi_D(z,s,\Delta\mathcal{X}_s) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma} \sum_{s'} \Psi(d_{ss'},\Delta d_s)^{\sigma}\kappa_{s'} - \Phi(\Delta\mathcal{X}_s;\Gamma_s) - \mathcal{F}^d_s. \tag{11}$$

We assume that taste for culturally distant goods takes the function form

$$\Psi(d_{ss'},\Delta d_s) = e^{-\hat{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s,k})^2\right)}, \tag{12}$$

where taste depends on the ex-post cultural distance between the firm and consumer $d^*_{ss',k} = d_{ss',k} - \Delta\mathcal{X}_{s,k}$.[16] The parameter $\hat{\beta}$ captures the taste elasticity with respect to caste distance ($d^2_{ss'}$). We use a first-order Taylor approximation of this function that gives us $\Psi(d_{ss'},\Delta d_{s'})^{\sigma} \approx$

[13] All results are valid for $\mathcal{N} \leq \mathcal{S}$.

[14] The vector $\Delta\mathcal{X}_s$ is always captured as the distance moved relative to firm owners' initial position.

[15] This assumption allows us to have the closed-form solution to the optimal caste location problem. This is isomorphic to the optimal location problem in the physical space.

[16] This functional form helps us to retain tractability of the model when we allow for both different taste and transportation costs in the next section.

$1 - \beta \left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right)$ with $\beta = \hat{\beta}\sigma$. This gives us a quadratic taste function that is reminiscent of Hotelling (1929) quadratic cost functions. The advantage of the quadratic taste function is that it gives us the closed-form solution to the optimal ethnic distance problem. Together, with this, we assume that the cost of moving in caste space is quadratic in ethnic distance and is given by

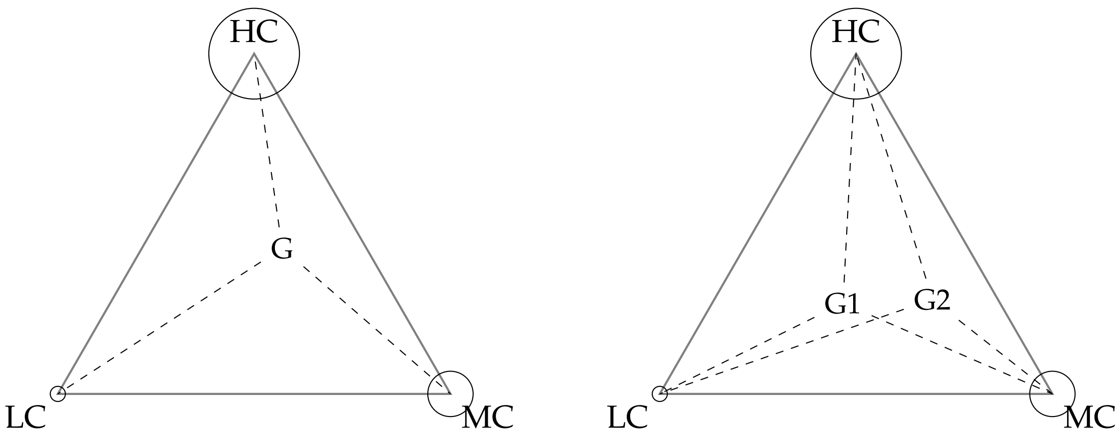
$$\Phi(\Delta\mathcal{X}_s; \Gamma_s) = \sum_{k\in\mathcal{N}} \gamma_{s,k} \Delta\mathcal{X}_{s,k}^2; \tag{13}$$

where $\gamma_{s,k}$ is the cost of moving along the dimension k for the firm owner of caste s.[17,18] We assume that these costs are paid in the wages of the labor that belongs to the castes towards which the firm owner wants to move closer. Using Equation A.4, we can rewrite profits as

$$\Pi_D(z, s, \Delta\mathcal{X}_s) = B_s Y \left(1 - \sum_{s'}\sum_{k} \lambda_{ss',k}(d_{ss',k} - \Delta\mathcal{X}_{s,k})^2\right) - \mathcal{F}_s^d, \tag{14}$$

where $B_s = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma}$ and $Y = \sum_s \kappa_s$, and $\lambda_{ss',k} = \frac{\beta\kappa_{s'} + B_s^{-1}\gamma_{s,k}\mathbb{1}_{s=s'}}{Y}$.

Figure 1: Optimal Ethnic Distance for Firms



(a) No Adjustment Costs

(b) Adjustment Costs

Note. In both figures, we consider a case with three castes (LC, MC, and HC) that are represented in a two-dimensional Euclidean space ($\mathcal{R}_+^3$). The size of the circles at the edges represents the relative size of each caste. Figure 1a shows the case where all firms face no costs to locate in caste space. All firms choose the same identity represented by point G. Therefore, there are no caste-specific differences across firms. The firms are closer to MC and HC as they offer access to larger markets even though that means that firms lose demand from LC consumers. Figure 1b show the case where the cost of moving in caste space is positive. Firm G2 belongs to the firm owner that was born as MC. It chooses position G2 which is closer to HC consumers to take advantage of their size. As it is costly to move closer to HC, its absolute advantage is lower relative to the case shown in Figure 1a. Similarly, firm G1 belongs to the firm owner that was born as LC. Her firm chooses position G1 which is closer to HC and MC consumers to take advantage of their size.

[17] The moving costs can differ for each caste in each direction. For instance, LC may find it easier to move closer to MC than HC. However, for our baseline calibration, the moving costs are the same across all castes and all dimensions.

[18] We are abstracting from a worker-side taste factor in the labor market.

Proposition 1. *The optimal ethnic distance for a firm is a relative size-based weighted average of the distance between the caste of the firm owner and the caste of the consumer and is given by*

$$\Delta\mathcal{X}_{s,k} = \frac{\sum_s \lambda_{ss'k} d_{ss',k}}{\sum_s \lambda_{ss'k}}, \quad \forall \quad k \in \mathcal{N}. \tag{15}$$

Proof. See Proof in Appendix A. ■

The above proposition states that firms move closer to the caste with the biggest size by more hiring workers. The distance move decreases with the cost of hiring workers from the target caste.

2.4 Taste Shifters and Transportation Costs

In the previous section, we only focused on taste-specific demand shifters. However, this may be confounded by the transportation costs if consumers of different castes reside in different locations (segregated regions). Under such conditions, the trade across castes will be affected by both tastes and transportation costs.[19] Here, we derive the solution to the optimal caste identity under both of these distortions. We assume that physical distance (that captures transportation costs) and caste distance (that captures taste) are perfectly correlated. This is a reasonable assumption since LC communities tend to be segregated.[20] Therefore firm owners from LC communities may have a need to include transportation costs when selling to consumers of high castes. We follow the trade literature and assume transportation costs as an iceberg cost. We assume that there is no transportation cost to sell to your own caste (as firms and consumers are in the vicinity of each other). In this case, the price of a product, which depends on to whom a firm is selling, is given by

$$p(z, s', s) = \frac{\sigma}{\sigma - 1} \mathcal{C}_s \tau_{s's} = \tau_{s's} p(z, s', s), \tag{16}$$

such that price of selling to a caste s' is $\tau_{s's}$ times the price of selling to own caste.

We assume that when firms change their ethnic identity, it also translates into a change in the physical distance and that they are perfectly correlated.[21] Further, we assume that transportation costs are quadratic in distance $\tau(d_{ss'}, \Delta d_s) = e^{\gamma\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s,k})^2\right)}$, where γ disciplines the rate at which the costs increase. The overall trade barrier for a firm can be summarized as composite $\Lambda(d_{ss'}, \Delta d_s) = \tau(d_{ss'}, \Delta d_s)^{1-\sigma} \Psi(d_{ss'}, \Delta d_s)^{\sigma}$. We use the first-order Taylor approximation of $\Lambda(d_{ss'}, \Delta d_{s'}) = 1 - \tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right)$, where $\tilde{\beta} = \hat{\beta}\sigma + \gamma(\sigma - 1)$

[19] We assume that the transportation costs are paid by the consumer. This is isomorphic to assume that consumers face higher search costs in finding the products of firms that are culturally distant and these costs raise the effective price per unit paid by the consumer.

[20] See, for instance, Bharathi, Malghan, Mishra and Rahman (2021) for recent evidence.

[21] For example, opening a branch closer to the neighbourhoods of other castes reduces transportation costs and also increases the perceived quality of the product.

that captures the strength of the cross-caste trade barriers. Under these assumptions, the solution is similar to Equation A.10, with the only difference being that the $\tilde{\beta}$, which captures the rate at which trade between two castes declines with ethnic and physical distance, is a combination of taste parameter β and transportation cost parameter γ. We provide more details in Appendix A.[22]

Lemma 1. *Let us define assimilated regions where firms face similar transportation costs to sell to consumers of different castes or $\tau_{ss'} = \tau = 1$. In such regions, $\tilde{\beta} \rightarrow \beta$, and only taste derives the difference in firm outcomes.*

Proposition 2. *Under endogenous cultural proximity, individual partial elasticity of firms' revenues with respect to the destination caste s to Income shocks I_s (all else constant) is given by*

$$\frac{\partial \log r(z,s,s^{'})}{\partial \log I_s} = \underbrace{\frac{\partial \log \kappa_s}{\partial \log I_s}}_{\textit{Size effect}} \underbrace{+2\tilde{\beta} \sum_{k \in \mathcal{K}} d^*_{ss',k} \frac{\partial \Delta \mathcal{X}_{s',k}}{\partial \log I_s}}_{\Delta \textit{ Optimal Distance}} \tag{17}$$

The variables Δ in optimal distance to the caste s in defined in the appendix. The overall revenue firms' elasticity is then given by

$$\frac{\partial \log R(z,s')}{\partial \log I_s} = \frac{\partial \log r(z,s,s^{'})}{\partial \log I_s} A_{ss'}, \tag{18}$$

where $A_{ss'}$ is the share of revenues for the firm owner of caste s' that is coming from consumers of caste s.

Proof. See Proof in Appendix A. ■

In Proposition 2, we provide the expression for the revenue elasticity to a demand shock to caste s'. There are two effects, the direct effect and the indirect effect. The former comes from the increase in the size of the caste and thus the demand for the firm's product. As consumer gets richer, they demand more on differentiated varieties, thus the total expenditure on differentiated varieties increases.[23] Second, the indirect effect captures the fact that firms change their cultural proximity to caste s, when its income increases. It is easy to see that firm owners selling to own caste consumer, this effect is positive, that further increases the caste-specific revenues elasticity. Our framework nests the standard trade model, where micro trade elasticity is given by the standard term. Under fully exogenous cultural proximity with $d_{s's'} = 0$ and $d_{ss'} > 0$, the revenue elasticity to demand shocks is reduced to only the direct effect.

[22] We abstract away from congestion and agglomeration forces for tractability.

[23] As income increases, there are more varieties available, thus the price index decreases. Under the exogenous distance benchmark, the net demand effect κ_s is always positive and the elasticity is similar to the one shown in Chaney (2008), with a small modification due to the presence of non-homotheticities in ou framework.

The elasticity of total firm-level revenue to an income shock is a product of caste-specific revenue elasticity and the share of revenue attributed to that specific caste. This implies that firms that sell relatively more to the caste that experiences an income shock also witness a higher increase in overall revenues. Next, we are interested in how these elasticities respond to the changes in the trade resistance parameter $\tilde{\beta}$.

Proposition 3. *Assume that the adjustment costs (Γ_s) are proportional to the trade-resistance parameter $\tilde{\beta}$. The elasticity of revenue shares to the trade resistance parameter $\tilde{\beta}$ (with constant σ) is given by*

$$\frac{\partial A_{ss'}}{\partial \tilde{\beta}} = A_{ss'} \sum_{s''} A_{s''s'} \left((d^*_{s''s'})^2 - (d^*_{ss'})^2 \right) \tag{19}$$

Proof. See Proof in Appendix A. ■

Corollary 1. *If the cost of adjusting caste proximity is high enough, the share of revenues coming from own caste consumers is increasing in $\tilde{\beta}$. Thus, the elasticity of the total firm revenues w.r.t. to income shock to caste s' of the owner with caste s' is higher than the revenue elasticity of the firm owner of any other caste.*

2.4.1 Firm Entry and Exit

To produce in the differentiated good sector, firms must pay a fixed entry cost, which is thereafter sunk. In equilibrium, with positive production of differentiated goods by firms of each caste, we require that the expected value of entry be equal to the sunk cost of entry f_e, which is paid in terms of labor. Firms draw their quality z from an exogenous distribution with CDF $G(z)$ and they enter the market if the profits are positive. If they do, they receive profits π every period they produce. Moreover, firms are risk-neutral and face an exogenous probability of exit δ. The free entry condition implies that

$$\int_z \max \left[0, \frac{\pi(z,s)}{\delta} \right] = w_s f_e, \quad \forall \quad s \in \mathcal{S}. \tag{20}$$

As wages across castes are the same, the expected profits before entry are also equalized across castes. In equilibrium, exiting firms are replaced by new entrants such that the firm distribution remains stationary. We denote by $\{L^e_s\}_{s\in\mathcal{S}}$ the mass of workers used for entry and by $\{M^e_s\}_{s\in\mathcal{S}}$ the mass of potential entrants each period. Let us denote the probability of survival of entrants by ω^e_s. In a steady state, the mass of operating firms M_s remains constant for each caste, such that $\delta M_s = \omega^e_s M^e_s$.

2.5 Equilibrium

Given the exogenous quality distribution $G(z)$, that is constant across castes, the equilibrium in this economy is a set of prices that includes wages for labor $\{w_s\}_{s\in\mathcal{S}}$, prices for each variety $\{p(z,s,s')\}$, and price for homogeneous good P_H, and consumption quantities $(C_H, \{c(z(\omega),s,s')\})$, output quantities $(y_H, \ell(z,s,s'), y(z,s,s'))$, and mass of active firms $\{M_s\}_{s\in S}$, mass of entrants $\{M_s^e\}_{s\in S}$ and labor used for entry $\{L_s^e\}_{s\in S}$.

1. Households maximize utility according to (1).
2. Producers of differentiated varieties maximize profits and charge the constant markup price given by Equation (8).
3. Product market clear for the homogeneous good and for each of differentiated goods.
4. The free entry condition holds.
5. Labor market clears for each caste.

2.6 Taking Stock

We have developed a theory to link the taste for identity in consumer demand to firms' incentive to trade across castes and hence the process of firm creation, the equilibrium firm size distribution, and the caste composition of the workforce. At the heart of our model is the insight that a taste for identity in consumer demand endogenously discourages firms from trading goods across castes and thus reduces their incentive to hire employees from diverse groups.

These trends align qualitatively with established observations about firm dynamics in India and other emerging markets where group boundaries are strong and persistent, characterized by the prevalence of small, non-expanding businesses, an abundance of subsistence producers, and the homophily in employee hiring. The surplus of small firms in emerging countries may not solely be attributed to the supply-side obstacles these firms encounter; it could also stem from firms facing a limited market to sell their high-quality products. In the following sections of this paper, we will assess the extent to which this mechanism can provide a quantitative explanation for the disparities in the distribution of firm sizes, and aggregate homophily in the economy. To do this, we need to estimate empirical elasticities that are crucial to identity model parameters. We first start by providing the details on institutional background.

3 Institutional Background

The caste system in India is a form of social stratification that historically divided people into rigid hierarchical groups based on their occupation. For centuries, caste has been assigned to individuals at birth, irrespective of their parents' current occupation. It is propagated through endogamy, and has dictated customary social interaction and exclusionary practices.[24] In descending order of dominance, the hierarchy dictated by the system, based on historical occupation, is as follows: Brahmins (priests and teachers), Kshatriyas (rulers and soldiers), Vaishyas (merchants and traders), and the Sudras (labourers and artisans). Further, two additional groups fall outside and below the caste system. The first one embodies the group of people traditionally known as Dalits.[25] The second group of people is known as the Scheduled Tribes.

For the remainder of the paper, we focus on a broader classification of the caste system, following Indian administrative practices. This choice is partly driven by the data constraints. Historically disadvantaged individuals belonging to the lower categories of the hierarchy are denoted by "LC," which includes the Schedules Castes and Scheduled Tribes; middle-category individuals are denoted by "MC," which includes the Sudras (also known as Other Backward Castes, OBC, falling between the traditional dominant upper and dominated lower caste-categories), and the individuals belonging to the historically privileged highest categories of caste hierarchy are denoted by "HC."

4 Data and Measurement

We integrate data from diverse sources to empirically examine the relationship between the caste composition of employees and the local distribution of demand. Our comprehensive dataset comprises information on the caste affiliations of both firm owners and employees across Indian districts. The dataset encompasses details on consumption patterns, demographics, and firm production.

4.1 Micro Small and Medium Enterprise (MSME) Data

The MSME data set is a nationally representative sample of micro, small and medium-sized firms in India for the financial year 2006-07 (recently used in Goraya, 2023). It consists of registered and unregistered firms. This classification is defined under the Factories Act 1948. A typical registered firm is larger and less likely to be a subsistence enterprise. In particular, the data set provides location, industry-product classification, balance sheet variables, and especially, the caste and gender of the firm owner and employees.[26] In total, we have approximately 596 thousand firms employing 3.3 million employees in 200 sectors producing 5000 distinct products and services. We provide the distribution of employees/revenues in

[24] Bidner and Eswaran (2015) describes the caste system as a 3,500 year old system. See, for instance, Deshpande (2010) for a discussion on the history of the caste system.

[25] In the Indian constitution, Dalits have fallen under the category of Scheduled Castes since 1947. Scheduled Castes is an officially designated group of historically disadvantaged people.

[26] The main advantage of using this data is the availability of the caste of the enterprise owner and employees. The MSME data set omits large enterprises (above a certain threshold of capital stock) unlike the ASI-NSS or the CMIE Prowess data sets. However, the latter data sets are not suitable for our analysis as they lack information on employee-employer castes.

All firms	Mean	Median	p5	p95	N
Emp. All	5.5	2	1	15	595964
Emp. LC	1.1	0	0	4	595964
Emp. MC	2.0	1	0	5	595964
Emp. HC	2.4	1	0	6	595964
Emp. OC (%)	75	100	0	100	595964
Revenue (10^4)	372	14.5	2.9	520	595964
Materials (10^4)	245	4.8	0.28	312	595964

Table 1: Firm Size Distribution

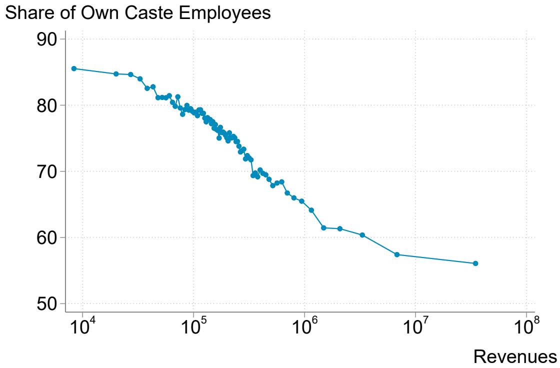


Figure 2: Own-Caste Employee Share

Notes: Table 1 presents the firm size distribution of the sample. Emp. All counts total employees in a firm; Emp. LC, Emp. MC, Emp. HC counts LC, MC and HC employees within a firm. Emp. OC presents the share of Own-caste workers (workers that belong to the caste of the employer). p5 and p95 are 5^{th} and 95^{th} percentile of the distribution, and N is the number of firms. Figure 2 presents a bin-scatter plot; the x-axis is the total revenues, and the y-axis is the share of Own-caste workers. We control for caste, district and 4-digit fixed effects. Sampling weights are applied.

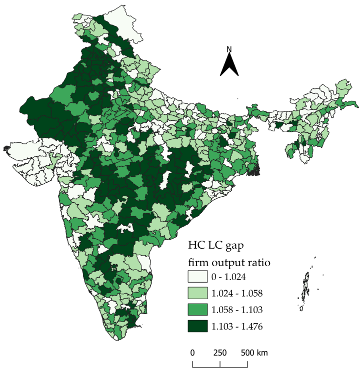
Table 1. Figure 3a plots the spatial variation in this asymmetry in entrepreneurial variables across caste, by plotting the absolute difference in the log of gross output (revenues) between HC and LC firms. The plot shows a lot of variation in this asymmetry, across districts in India that we will exploit in our empirical analysis.

Caste composition of employees: The novel part of the data is the employer-employee caste linkages. We establish two main facts: first, there is homophily in employee hiring; 75% of the workers belong to the caste of the firm owner on average (Table 1). Second, homophily declines with firm size, see Figure 2. This is not driven by particular castes, districts or sectors (see Figure B.2 in appendix). To the best of our knowledge, this is the first paper to document the segregation of firm-level employment by caste in India.

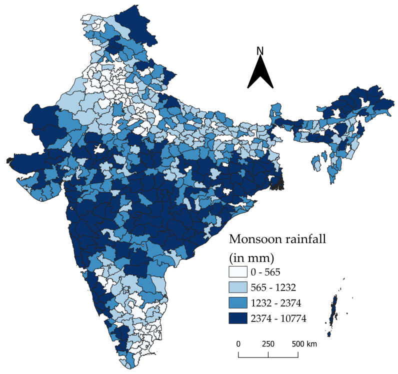
Balanced panel 2004-2006-07: The MSME data does provide retrospective information on *revenues* and *material purchased* for the firms that survive up to 2006-07. This allows us to construct a balanced panel of MSMEs for the three years, 2004-05 to 2006-07. This panel of firms allows us to use the time and regional variation in rainfall and revenues across castes.

Prices and quantities: The MSME data provides information on revenues and quantities for three main *products* and three main *input materials* for the cross-section of firms during 2006-07 (see Appendix B.1 for more details). We compute average product prices by dividing product revenues by quantities. The product price data plays an important role in laying out differences between taste for identity and quality. We use this data to distinguish between the price and quantity effect of demand shocks. Further, the price data is used to separate standardized products from customized products. The main assumption is that the low price dispersion is linked to standardized products. Finally, we will exploit price data to infer product quality differences across castes.

Figure 3: Spatial Variation in Monsoon Rainfall and Firm Size, at the District-Level in India (2006-07)



(a) Spatial variation in output-difference between HC and LC firms



(b) Spatial variation in monsoon rainfall

Notes: Figure a uses MSME 2006-07 data to plot the absolute difference in log(gross output value) between firms owned by members historically classified as high-caste and firms owned by members historically classified as low-caste. Figure b uses TRMM 2006-07 data to plot the total monsoon rainfall received by Indian districts, in millimeters of rainfall.

4.2 Consumption Data

We use data from the National Sample Survey (NSS) Household Consumer Expenditure survey on households, their consumption, and their demographics, at the district level. We use four schedules spanning the years 2003-2004 to 2007-08. Specifically, the analysis includes the NSS waves 60, 61, 62, and 63. The survey includes questions about the activities of individuals during the most recent seven days. We use the *monthly per capita expenditure* (MPCE) as the measure of consumption. This is computed as total monthly expenditure divided by household size. We consider both total MPCE and MPCE in different consumption categories. In particular, we classify expenditure into four categories - all, manufacturing (clothes and footwear), services, and durable goods. To obtain real consumption, we divide nominal consumption by the state-level "Consumer Price Index for Agricultural Labourers", published by the Government of India. The descriptive statistics are provided in Table B.1. Panel C in Table B.1 shows that, on average, HC households' monthly per-capita expenditure (1234.5 Indian rupees) is nearly double that of LC households (627.1 Indian rupees).

4.3 Rainfall Data

We use data from the Tropical Rainfall Measuring Mission (TRMM), developed by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace and Exploration Agency (JAXA). Figure 3b plots the spatial variation in rainfall, measured in millimeters of monsoon rainfall. The plot shows a lot of variation across districts in India.

5 Empirical Strategy

Motivated by the model predictions documented in Section 2, our empirical specification seeks to estimate the effects of caste-linkages on local firms' outcomes in India. We define a local economy as the economy of an Indian district.[27] We posit that in the presence of taste for identity in consumer demand, caste-linkages determine firms' demand. We shed light on this linkage by using an exogenous shift in local demand for LC households, due to higher rainfall, and observe its effects on firms owned by different caste-groups across different sectors.[28]

5.1 Empirical Analysis

5.1.1 Effect of rainfall on household consumption

We hypothesise that the asymmetric benefits of higher rainfall especially increase LC household consumption, through the effect of rainfall on agricultural wages (see Appendix B.4 for evidence on wages). Using data on consumption patterns across product categories such as Manufacturing, Services, and Durables, we would also be able to investigate the effect of higher rainfall as a demand-shifter across non-agricultural product categories. To do this, we estimate the following equation:

$$log(c_{ht}) \;=\; \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot rainfall_{dt} \times caste_i + \delta_{d,t} + \delta_c + \epsilon_{ht}, \tag{21}$$

where h denotes household. The regression includes caste and the district $\times$ year fixed effects. The coefficient $\beta_{2,LC}$ gives the elasticity of LC households' consumption to rainfall. According to our hypothesis, higher rainfall increases their consumption more ($\beta_{2,LC} > 0$), relative to HC households.

5.1.2 Effect of rainfall on firms

Revenue Elasticity. Next, in establishing the link between LC households and LC firms, we check whether the above shift in LC households' demand particularly makes LC firms bigger. To do so, we estimate the following equation:

$$log(y_{ft}) \;=\; \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot rainfall_d \times caste_i + \delta_{d,t} + \delta_{s,t} + \delta_i + \epsilon_i, \tag{22}$$

where f denotes the firm and s denotes the industry. We estimate the effect of higher rainfall on firm-level (y): (1) revenue, and (2) material input. The regression includes cast and district $\times$ year fixed effects to control for any time-varying district-specific characteristics.

[27] Districts are administrative units within states and are analogous to counties in the US system.

[28] See Appendix B.4 for alternative measures of rainfall as a demand shifter.

Note that this fixed effect subsumes the average effect of rainfall as well, along with district-specific trends (e.g., migration). We rely on the residual variation, that is the asymmetry in economic conditions across caste and further, its asymmetric interaction with the levels of Monsoon-rainfall across districts, to obtain a plausibly causal interpretation. The regression also includes a 4-digit industry $\times$ year or a 5-digit product $\times$ year fixed effect to control for any market-specific characteristics common across the districts. The coefficient $\beta_{2,LC}$ provides the elasticity of LC firms' equilibrium outcomes to rainfall.

Workforce Caste Composition. To analyse the caste composition of the workforce, we use the cross-sectional data and run the following regression:

$$log(y_f) \;=\; \alpha + \beta_1 \cdot rainfall_d + \beta_{2,i} \cdot rainfall_d \times caste_i + \delta_d + \delta_s + \delta_i + \epsilon_i, \tag{23}$$

where firm-level outcome variables are caste-specific employee shares. The only difference relative to equation 22 is that we cannot have year-fixed effects in this specification. The coefficient $\beta_{2,LC}$ identifies the changes in caste-specific employee shares for LC firms relative to HC firms. Motivated by model predictions, we expect $\beta_{2,LC}$ to be positive when the outcome variable is LC and MC employee shares, whereas we expect $\beta_{2,LC}$ to be negative when the outcome variable is HC employee share.

However, we can not identify the response of firms owned by HC as that is absorbed by district fixed effects. To make progress on this front, we exploit variation in market competition. We hypothesise that, in the markets that are dominated by HC firms, the incentive for HC firms to hire LC workers is low as LC consumer has low-to-no outside option other than consuming goods produced by HC firms. We defined the market by district $\times$ sector. We label a market as $Competitive_{d,s}$ if the market share of HC firms is less than x percent, where $x \in \{50, 60, 70, 80, 90\}$. The market share can be defined in terms of the number of firms, revenues, or employment. We provide results with all possible combinations. Within the firms owned by HC, we run the following regression:

$$log(y_f) \;=\; \alpha + \beta_1 \cdot rainfall_d + \beta_2 \cdot rainfall_d \times Competitive_{d,s} + \delta_d + \delta_s + \epsilon_i, \tag{24}$$

where the firm-level outcome variable is the share of LC employees and we expect β_2 to be positive. Further, we exploit one more dimension of sectoral heterogeneity. We conjecture that the incentive to hire from the target consumer group should be higher in customer-facing or contact-intensive industries. In this case, we use the definition given by the Ministry of MSME for Repair & Maintenance industries, and Services. This is particularly useful in this context as the traditional view of considering services as more contact-intensive would fail to capture a large number of manufacturing industries that are also customer-facing (e.g., Tailoring). Within the firms owned by HC, we run the regression:

$$log(y_f) = \alpha + \beta_1 \cdot rainfall_d + \beta_2 \cdot rainfall_d \times Contact - Intensive_s + \delta_d + \delta_s + \epsilon_i. \quad (25)$$

where the firm-level outcome variable is the share of LC employees and we expect β_2 to be positive.

5.2 Empirical Results

5.2.1 Household consumption elasticity

Table 3 present the results from estimating Equation (21), and show the asymmetric effect of higher rainfall on Monthly Per Capita Expenditure (MPCE). LC households' consumption increases by 10%, for every 1% increase in rainfall relative HC households. We also find an increase in the MC households' MPCE. A substantial fraction of this increase is explained by an increase in spending on Services and Durables goods.

Table 2: Household Consumption Elasticity in Rural India

	Household Monthly Consumption per capita			
	(1) All	(2) Manufacturing	(3) Services	(4) Durables
log(rainfall) × MC	0.058***	0.042**	0.155***	0.113*
	(0.022)	(0.020)	(0.056)	(0.062)
log(rainfall) × LC	0.100***	0.048**	0.138***	0.280***
	(0.021)	(0.024)	(0.051)	(0.056)
Observations	154,241	153,586	132,873	126,125
R-squared	0.351	0.251	0.225	0.218
Controls$_{it}$	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓
Caste FE	✓	✓	✓	✓

Notes. The regressions are of household-level monthly per-capita expenditure (logarithmic) on rainfall (logarithmic) by caste. *Manufacturing* stands for consumption of clothing and footwear. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at district level in all regressions, *** p<0.01, ** p<0.05, * p<0.1.

5.2.2 Firm revenue elasticity

The above evidence shows that the higher rainfall induces an effect on the local economy with a shift in demand for goods and services, largely driven by LC households. We now evaluate the effect this has on firm outcomes, across different castes. Table 3 presents the results from estimating Equation (22), and shows the asymmetric effect of higher rainfall on firms' outcomes. We document an increase in the revenue of LC firms by 17.3%, relative to HC firms, for every 1% increase in rainfall. Consistent with our hypothesis, this observation suggests that LC firms gain the most due to a positive shift in demand from LC households,

Table 3: Firm Size Elasticity in rural India

	All Sectors				Downstream Sectors			
	(1) Revenues	(2) Inputs	(3) Revenues	(4) Inputs	(5) Revenues	(6) Inputs	(7) Revenues	(8) Inputs
log(rainfall) × MC	0.158***	0.191***	0.128***	0.149***	0.134**	0.154**	0.126***	0.149***
	(0.045)	(0.061)	(0.037)	(0.051)	(0.053)	(0.067)	(0.041)	(0.054)
log(rainfall) × LC	0.173***	0.226***	0.130***	0.172***	0.195***	0.232***	0.175***	0.215***
	(0.046)	(0.059)	(0.038)	(0.050)	(0.060)	(0.080)	(0.048)	(0.066)
Observations	1,468,668	1,457,139	1,465,826	1,454,276	610,699	605,585	610,348	605,226
R-squared	0.422	0.460	0.528	0.549	0.347	0.396	0.476	0.497
Caste FE	✓	✓	✓	✓	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Sector × Year	✓	✓			✓	✓		
Product × Year FE			✓	✓			✓	✓

Notes. The regressions are of firm-level variables (logarithmic) on rainfall (logarithmic) by caste. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at district level in all regressions, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

thus highlighting caste-linkages as a determinant of firm demand. The LC firms also witness a 22.5% rise in material input purchase, relative to HC firms, for every 1% increase in rainfall. These results remain robust to the inclusion of product × year fixed effects. Further, focusing on the downstream sectors such as services, and furniture manufacturing, among others, we find even higher LC's relative revenue elasticity.

5.2.3 Workforce Composition

To further understand the increase in the size of LC firms and the caste linkages in employment, we decompose the change in the labour composition of firms by caste. We provide three different forms of evidence to document the change in workforce composition: (1) rainfall triggered differences in LC and MC firms' workforce composition relative to HC firms, (2) rainfall triggered changes in HC firms' LC employee share in competitive markets relative to HC-dominated markets, (3) rainfall triggered changes in HC firms' LC employee share in contact-intensive industries.

Relative difference in Workforce Composition. First, we run the regressions in Equation 23 among the firms that hire HC employees and provide results in Table 4. We selected this sample because we want to focus on the intensive margin of hiring HC workers. It is evident from the results, that LC firms are larger in districts with higher rainfall, they have a lower share of HC workers, and they have a higher share of LC and MC workers (i.e., the caste groups that experienced a heightened demand). All these changes are measured relative to HC firms. This evidence supports the view that when LC and MC consumer demand is high then the incentive to hire workers from consumer groups is also high.

Table 4: Caste Composition of Employees Across Firms and Elasticity to Rainfall

	(1) Total Employees	(2) HC Share	(3) MC Share	(4) LC Share
log(rainfall) × MC	0.109***	-0.041***	0.036***	0.006*
	(0.028)	(0.010)	(0.007)	(0.004)
log(rainfall) × LC	0.145***	-0.040***	0.013**	0.027***
	(0.031)	(0.011)	(0.007)	(0.009)
Observations	263,051	263,051	263,051	263,051
R-squared	0.385	0.267	0.258	0.194
Caste FE	✓	✓	✓	✓
District FE	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓

Notes. The regressions are of using Equation (23) among the firms that hire HC employees. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Workforce composition of HC firms in Competitive markets. We run regressions using Equation 24 specifically within the HC firms. Market delineations were defined by district × sector. Figure 4a presents the findings using 3-digit sectors, while Figure 4b utilizes 4-digit sectors. The results underscore that in competitive markets, HC firms tend to hire more LC employees when the demand from the LC segment is elevated in comparison to markets dominated by HC firms.

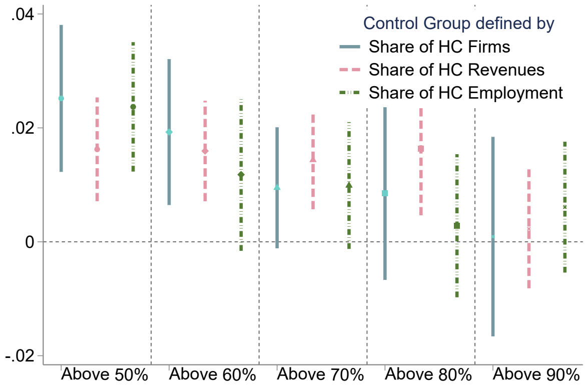
These results also emphasize that any changes in LC labor supply induced by rainfall are not the primary catalyst for the observed shift in the composition of HC firms' workforce. This is because rainfall-induced changes in LC labor supply may present at the district level, yet there is no apparent rationale for expecting these changes to correspond with variations in competition within specific industries.

Workforce composition in contact-intensive sectors. In our third string evidence, we run the regressions using Equation 25 within HC firms and provide results in Table 5. The results show that HC-owned firms hire more LC employees in contact-intensive industries in regions where LC demand is high. The main idea is that caste identity becomes more salient when there is customer-employee interaction, thus incentivizing HC-owned firms to hire more LC employees to cater to LC consumers. We use four different definitions of contact-intensive industries (list is provided in Appendix B.1) and results remain robust. All three pieces of evidence taken together highlight that firms capture diverse consumer markets by hiring employees from the target consumer caste.

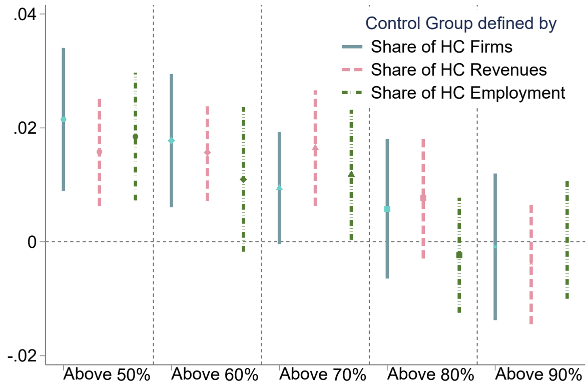
5.3 Robustness Checks

We have previously provided evidence of ethnic connections between businesses and consumers. Nonetheless, this observed overall separation could be attributed to the interplay

Figure 4: LC Employee Share Among HC-owned Firms in Competitive Markets



(a) Difference in LC worker Share: NIC 3



(b) Difference in LC worker Share: NIC 4

Notes. The figure presents the β_2 (y-axis) from regressions using Equation 24 among the HC firms. The dependent variable is LC employee share. We defined the market by district $\times$ sector. We label a market as $Competitive_{d,s}$ if the market share of HC firms is less than x percent, where $x \in \{50, 60, 70, 80, 90\}$. The market share is defined in terms of the number of firms (solid line), revenues (dashed line), or employment (dashed-dotted lne). Figure 4a presents the findings using 3-digit sectors, while Figure 4b utilizes 4-digit sectors. The band around the point estimate represents the 95% confidence interval.

of two underlying mechanisms: (a) taste for identity, where consumers prefer goods from their own group and (b) geographical distance, where consumers incur costs when accessing products located at a greater geographical distance. To separate these two forces, we employ village-level population data to assess the level of geographical segregation within each district. This data on caste population shares is sourced from the SHRUG database (Asher and Novosad, 2019). Our measure for geographical segregation quantifies the standard deviation of the local community (LC) population share within a district (see Figure B.3 in the appendix that shows the distribution for Tamil Nadu). We conduct a supplementary analysis by excluding observations from districts exhibiting high levels of geographical segregation. The results of this analysis closely mirror our baseline estimates, suggesting that geographical segregation may not be the primary driving force behind the observed patterns (see Table B.4).

We use the information on product prices and compare the firms that have similar prices within a product market. We find similar results to our baseline specification, highlighting the fact that taste for identity in consumption is not driven by physical distance or differences in observed quality (see Table B.11). In Table B.6, we show that results are not driven by small firms or by product markets that are dominated by LC entrepreneurs.

6 Model Quantification

In this section, we use our framework presented in Section 2 and the empirical estimates from Section 5.2 to quantify the effect of taste for identity in demand on cross-caste trade,

Table 5: LC Employee Share Among HC-owned Firms in Contact Intensive Industries

	(1) Share of LC Workers	(2) Share of LC Workers	(3) Share of LC Workers
log(rainfall) × Repair and Maintenance	0.013**		
	(0.005)		
log(rainfall) × Services	0.011*		
	(0.006)		
log(rainfall) × Contact-Intensive sectors		0.014***	
		(0.003)	
log(rainfall) × Contact-Intensive Services			0.009**
			(0.004)
Observations	218,410	218,410	218,410
R-squared	0.175	0.174	0.174
District FE	✓	✓	✓
Sector FE	✓	✓	✓

Notes. The regressions are of using Equation (25) among HC-owned firms. The Repair & Maintenance and service sector definitions are provided by the Ministry of MSME. The list of industries in Contact-Intensive sectors and Contact-Intensive services is provided in Appendix B.1. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

diversity in hiring and firm size. To do so, first, we need to calibrate the values for multiple parameters. Second, we provide policy simulations and aggregate cost/benefits using our benchmark calibration. We start by elaborating on the layout for our calibration strategy.

6.1 Calibration

The model requires us to provide values for thirteen parameters. We divide the model parameters into two major groups: Fixed and Fitted. The values for fixed parameters are either normalized to one or taken from the literature due to difficulties in calibrating them with existing data. Meanwhile, the fitted parameters are calibrated by matching certain moments in the data to their counterpart in the model. Next, we illustrate the fixed parametric values.

6.2 Fixed Parameters

In the model, we have allowed for heterogeneity across castes for production technology and cost structure, which we will shut down in this section. We will assume that all firms in the differentiated good sector use a production technology with labor from their caste for production. The productivity of the homogeneous good sector A_H is normalized to one so all wages are equal to one as well. Thus, the marginal cost of production is also equal to 1. The fixed cost of entry is normalized to one following Bai, Jin and Lu (2019).

For the moving costs in the ethnic space, we assume all costs are paid in labor. Further, we assume that costs are scaled by the quality parameter (that in turn pins down the firm size), such that the effective marginal benefit of moving is the same across firms. In particular, we assume $\gamma_{s,k} = \gamma B_s$, where $B_s = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} z^{\sigma}$ and $\gamma = 1$. This implies that, conditional on trading, the ethnic location of firms is independent of product quality or firm size. Thus, within firms of certain a group, all the variation in ethnic location and thus the ethnic composition of the workforce comes from trading and non-trading firms.[29] Further, following the previous literature (see, Broda and Weinstein, 2006), we set the elasticity of substitution to $\sigma = 5$.

6.2.1 Fitted Parameters.

There are three sets of parameters that need to be calibrated: household sector, firm dynamics, and cross-caste trading. First, we need the labor endowment for each caste, the share of homogeneous good sector a, and the taste elasticity $\beta \equiv \hat{\beta}\sigma$. There are parameters that are related to firm dynamics; probability of firm death δ, fixed cost of operating c_f, the sunk cost of entry c_e, and scale parameter of the Pareto distribution for firm productivity. Finally, there is a fixed cost of trading/hiring across castes c_x. We estimate c_x in terms of the fixed production costs $c_x = xc_f$.

Further, labor endowment for each caste, the share of homogeneous goods in consumption, and the probability of firm death are externally calibrated, see Panel B in Table 6. The rest of the parameters are internally calibrated (jointly) by matching specific moments as mentioned below.

6.2.2 Targeted Moments

Partial Equilibrium Revenue Elasticities. The main parameter of interest is the taste elasticity β. We use our reduced form estimates to infer the value of taste elasticity. In our empirical results, we find that a 1% increase in rainfall increases monthly consumption by 10% for LC consumers and 6% for MC consumers. This further led to an increase of 17% in yearly revenues. We will assume that this increase is uniformly spread across all months, which gives us a 1.3% monthly increase in revenues. In the model, we match this revenue increase in partial equilibrium, i.e., all else equal, a 10% percent increase in the market demand from LC consumers and a 6% increase in the market demand from MC consumers should increase LC firms' revenue by 1.3%. The model observes a monotonic relationship between the taste for identity β and revenue elasticity, see Figure 5a.

Firm Dynamics Moments. Two parameters need to be calibrated that are related to firm dynamics: fixed cost of operating c_f and scale parameter of the Pareto distribution for firm productivity. We match the percentage of revenues produced by the bottom 50 percent of

[29] This is a simplification assumption that reduces the number of unknown parameters in the model. In principle, we could match the profile of firm size and ethnic concentration of employees to back out the cost profile. However, there could be many other forces that are missing from the model but could potentially derive the ethnic concentration of employees within a firm. For this reason, we restrict ourselves to this calibration strategy.

firms and the top 10 percent of firms. In the model, an increase in fixed costs increases the share of output produced by the bottom half of the revenue distribution, see Figure 5b, whereas a decrease in the scale parameters η, makes the tail of the productivity distribution thicker, increases the share of the top 10 percent, see Figure 5d.

Trade Moment. We infer the fixed cost of trading by matching the percentage of firms that hire outside their caste. In our model, the necessity for firms to hire employees from the caste they are selling to implies an equivalence between the share of firms hiring and trading outside one's caste. While this assumption may not precisely mirror data, we find empirical support, with the share of firms hiring outside one's caste approximating 38% in our data, closely mirroring the probability of trading outside one's caste documented in a recent paper by Boken, Gadenne, Nandi and Santamaria (2022).[30] According to their estimates firms are twice as likely to trade within their caste relative to trading with outside castes. In the model, a higher cost of trading results in a lower share of firms hiring workers outside their caste, see Figure 5b.

Table 6: Fitted Model Parameters and Targetted Moments

Parameter	Value	Description	Moment	Model	Data
Panel A. Internal Calibration					
Household Sector					
β	0.308	Taste elasticity	Revenue Elasticity	0.013	0.013
Firm Dynamics					
c_f	0.02	Fixed operating cost	Sales share: bottom 50%	1.3%	1.3%
η	3.20	Scale parameter: Pareto distribution	Sales share: Top 10%	86.3%	90.1%
Trading					
c_x	$3.05\times c_f$	Fixed trading cost	Share: hiring out-caste worker	38.1%	38.2%
Panel B. External Calibration					
$\{L_{LC}, L_{MC}, L_{HC}\}$	{0.80,0.90,1.00}	Household income shares			
a	0.30	Share of homogeneous good			
δ	0.088	Probability of firm exit			

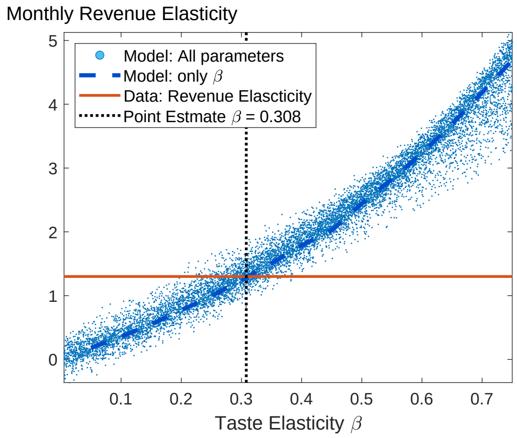
Notes. Panel A lists the internally calibrated parameters and corresponding targetted moments. The moments are computed using the MSME data and the monthly revenue elasticity is computed by dividing 17.3% (the yearly relative revenue increase for LC firms as in Table 3) by 12 (number of months). The Sales share: Top 10% and Sales share: bottom 50% computes the revenue share of the top 10% percent of the firms and bottom 50% percent of the firms. The Share: hiring out-caste workers computes the share of firms hiring workers outside their castes. Panel B presents parameters that are externally set: household income share and share of expenditure on homogeneous goods are computed from the NSS consumption survey. The probability of firm exit is taken from Hsieh and Klenow (2014).

6.3 Results: Taste for Identity, Firm Size and Homophily in Hiring

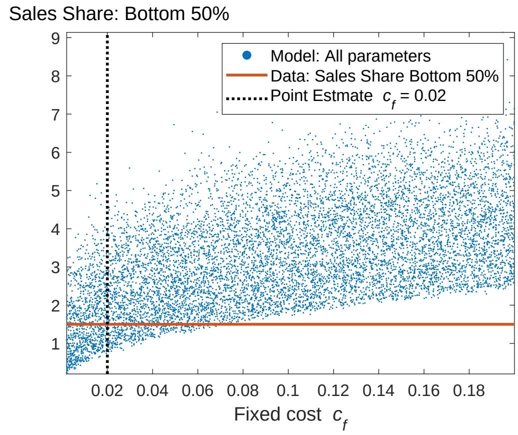
The estimated parameters and the corresponding moments are provided in Table 6. The value for taste for identity parameter is estimated to be 0.308, the fixed cost of production is

[30]Their dataset is based on firm-to-firm trades within West Bengal.

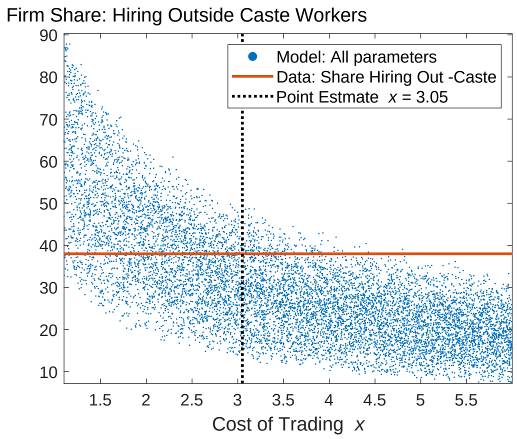
Figure 5: Identification of Taste Elasticity, Cost of Production & Trading and Productivity Distribution.



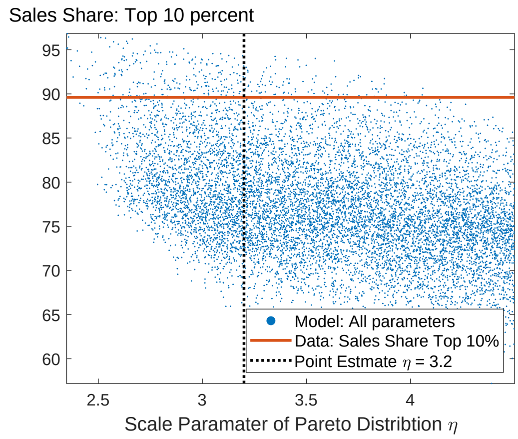
(a) Taste For Identity



(b) Fixed Costs



(c) Cost of Trading



(d) Scale Parameter

Notes: This figure provides the relationship between parameters and targeted moments. Blue dots represent model simulations sequencing parameter vectors $\Theta = \{\beta, \eta, c_f, x\}$ using a Sobol sequence in the 4-dimensional terrasect $[0.005, 0.15] \times [2.0, 4.5] \times [0.002, 0.20] \times [1.1, 6.0]$. The dashed dark blue line presents a moment in the model when only the parameter on the x-axis varies, holding other parameters at their point estimate. The horizontal solid orange line presents a moment in the data. The vertical dashed black line shows the parameter estimate.

equal to 0.02, the cost of trading is 3.05 times the fixed cost, and the scale parameter of the Pareto distribution is estimated at 3.20.

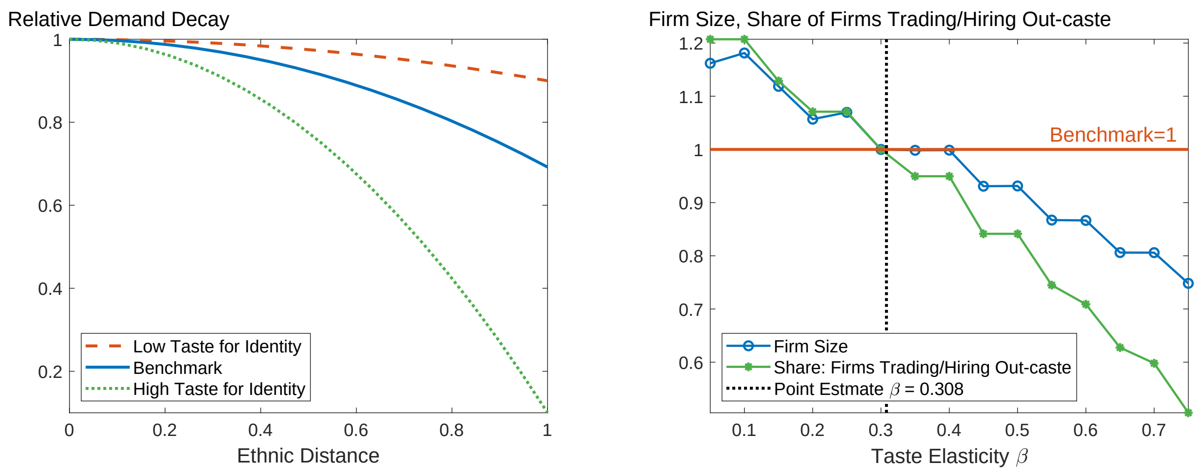
The model finds a substantial taste for identity in the economy with demand decaying at a rate of 30.8% over the ethnic distance between caste. In Figure 6a, we compare the demand decay rate for different levels of taste for identity parameter β and show the relationship between demand decay and taste for identity is convex.

Within the benchmark economy, the distribution of firms among different castes exhibits a magnitude surpassing proportionality to their respective labor contributions. This finding draws parallels with the work of Helpman, Melitz and Yeaple (2004) in the field of exports. Consequently, the high-caste firms assert control over the majority of the market share within the differentiated goods sector, while the low-caste firms entities specialize in the production of homogeneous goods. Notably, the economically disadvantaged LC group experiences a substantial dip in real income compared to the affluent HC group, surpassing the differences in labor endowment. This preference for identity acts as an implicit trade friction, resulting in a reduction of real income for all caste groups.

In Figure 6b, we delineate the aggregate ramifications of the taste for identity in consumer utility. Initially, this inclination discourages firms from venturing into markets that are ethnically distant, leading to a constrained firm boundary that remains predominantly local. This limitation stems from two primary factors: firstly, firms incur a fixed cost when hiring or accessing an ethnically distant market. Consequently, only a select few firms producing high-quality products find it economically justifiable to bear this fixed cost and engage in selling to diverse caste groups. Moreover, as the taste for identity intensifies, the benefits diminish further, resulting in an even smaller fraction of firms participating in cross-caste hiring and trade, as illustrated in Figure 6b, showcasing the monotonic and decreasing relationship between the share of firms engaging in cross-caste hiring and the taste for identity parameter β.

Secondly, firms face a dilemma when attempting to sell to multiple markets, as doing so adversely impacts the demand from their group – a trade-off commonly absent in the conventional models of international trade (see, Melitz (2003)). In this context, when firms hire workers from other castes or sell to other castes, the demand from their group decreases. This critical trade-off diminishes the incentive for diversification. Both the extensive margin of selling and the intensive margin of improving taste contribute to smaller firm sizes. This outcome provides a demand-driven explanation for the observed phenomenon of low firm size in India, as highlighted by Hsieh and Klenow (2014).

Figure 6: Taste for Identity, Homophily in Hiring and Firm Size



(a) Demand Decay and Taste for Identity

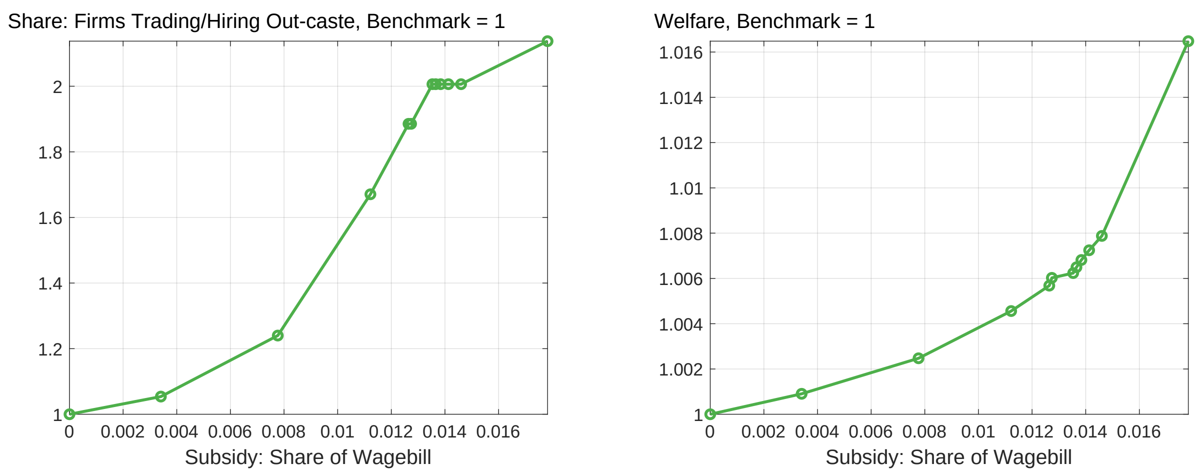
(b) Firm Size and Trading/Hiring Decision

Notes: This figure plots the counterfactual aggregate outcomes under different values of parameters β that measure taste of identity in demand. Figure 6a plots the demand decay over the ethnic distance between the firm owner and consumer and Figure 6b computes the change in average firm size and share of firm trading/hiring outside their caste relative to the benchmark calibration (with $\beta = 0.308$), the benchmark values are normalized to one.

6.4 Policy Analysis: Should We Subsidise Diversity in Hiring?

Utilizing our calibrated model, we pose a straightforward question: Does the provision of subsidies for cross-caste hiring contribute to overall welfare improvement? To answer this question, we solve the model for different levels of γ and x, that capture the marginal cost and fixed cost of of cross-caste hiring, respectively. The subsidy is formally defined as the difference between the firms' wage bill allocated to cross-caste hiring under the prevailing cost structure and the corresponding expenditure under the cost structure of the benchmark economy. This subsidy is revenue-neutral in the following sense. In this model, total nominal revenues remain constant, i.e., the sum of labor endowment of each caste times their wage. Thus, the reduction in hiring costs is a transfer from households to firms that in turn is paid back to the household in terms of wages. Across our policy experiments, the subsidy varies between 0 and 2 percent of the total wagebill of the differentiated good sector.

Figure 7: Cost of Cross-caste Hiring, Homophily and Real Income



(a) Share of Firms Trading/Hiring Out-caste

(b) Welfare: Real Income

Notes: This figure plots the counterfactual aggregate outcomes under different values of subsidy to cross-caste hiring. Figure 7a plots the share of firms trading/hiring outside their caste and Figure 7b computes the real income for relative to the benchmark calibration, the benchmark values are normalized to one.

The reduced costs act as an incentive for firms to engage in selling to ethnically distant castes. This implies that economic barriers or challenges related to cross-caste trade are being addressed by the subsidy. As this expansion entails hiring more from other castes, the subsidy contributes to an overall reduction in homophily across the economy, see Figure 7a. The number of firms selling across castes multiplies exponentially with subsidy and this is the primary driver of the next set of results.

The entry of firms into diverse markets leads to heightened competition among firms as the demand for labor increases. This results in an increase in number of firms serving each ethnic market. Consequently, the average firm size increases. Further, there is a reallocation of economic activity from LC firms towards HC firms. Thus, the competition erodes the market share of firms of castes that have small size of labor endowment, i.e., LC in our framework, and an increase in the market share of firms that belong to large groups, i.e., HC. In this sense, the inequality of firm outcomes increases across castes.

However, even though the number of LC-owned firms declined, the number of varieties available to LC consumers and all other consumer groups increased as firms could access multiple markets, leading to an overall improvement in welfare as measured by real income for all groups, see Figure 7b.

7 Conclusion

In this paper, we show that, in a context where demand is driven by identity, workforce diversity acts as a tool to gain access to diverse consumer markets. We present a model where consumers care about the observed quality of goods and the identity of the producers. Both of these attributes create the perceived quality of the goods. Firms can sell goods to different markets, but it is more difficult to sell to ethnically distant markets. The ethnic distance between castes is exogenous and defined by existing norms. The firms, however, know about preferences and actively pursue to increase the perceived quality of their product. One way to do it is to hire workers from the target consumer group. The main result of the paper is to show that the optimal hiring of outside group employees is proportional to their market demand and the cost of hiring.

We apply this theory to the caste norms in rural India. We use novel data on employer-employee caste linkages and exogenous changes in consumer demand to provide causal evidence on the main mechanism. We show that this shock, one, resulted in a much higher increase in the revenue of the same caste, and second, the workforce composition of firms shifted towards this consumer caste. Finally, We use these micro-elasticities together with the model to show that the taste of identity in demand is substantial. This makes demand ethnically local and generates small firm size and more homophilic hiring. We use this model to understand the role of policy in this environment. The model highlights that a subsidy to diversify in hiring – cross-caste hiring in this context – promotes firm entry into ethnically distant markets. This creates less homophily in hiring and increases welfare for consumers.

References

Adhvaryu, Achyuta, Amalavoyal V Chari, and Siddharth Sharma, "Firing costs and flexibility: evidence from firms' employment responses to shocks in India," *Review of Economics and Statistics*, 2013, *95* (3), 725–740.

Afridi, Farzana, Amrita Dhillon, and Swati Sharma, "The ties that bind us: social networks and productivity in the factory," 2020.

Agte, Patrick and Arielle Bernhardt, "The Economics of Caste Norms: Purity, Status, and Women's Work in India," 2023.

Aker, Jenny C, Michael W Klein, Stephen A O'Connell, and Muzhe Yang, "Borders, ethnicity and trade," *Journal of Development Economics*, 2014, *107*, 1–16.

Albornoz, Facundo, Héctor F Calvo Pardo, Gregory Corcos, and Emanuel Ornelas, "Sequential exporting," *Journal of International Economics*, 2012, *88* (1), 17–31.

Alesina, Alberto and Eliana La Ferrara, "Participation in heterogeneous communities," *The quarterly journal of economics*, 2000, *115* (3), 847–904.

_ and _ , "Ethnic diversity and economic performance," *Journal of economic literature*, 2005, *43* (3), 762–800.

_ , Reza Baqir, and William Easterly, "Public goods and ethnic divisions," *The Quarterly journal of economics*, 1999, *114* (4), 1243–1284.

Alfaro-Urena, Alonso, Juanma Castro-Vincenzi, Sebastián Fanelli, and Eduardo Morales, "Firm export dynamics in interdependent markets," Technical Report, National Bureau of Economic Research 2023.

Allen, Treb, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

Alsan, Marcella, Owen Garrick, and Grant Graziani, "Does diversity matter for health? Experimental evidence from Oakland," *American Economic Review*, 2019, *109* (12), 4071–4111.

Anderson, Siwan, "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics*, 2011, *3* (1), 239–263.

Asher, Sam and Paul Novosad, "Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG)," 2019.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2019, *17* (6), 1881–1940.

Atkin, David, Amit K Khandelwal, and Adam Osman, "Exporting and firm performance: Evidence from a randomized experiment," *The Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

Bai, Yan, Keyu Jin, and Dan Lu, "Misallocation under trade liberalization," Technical Report, National Bureau of Economic Research 2019.

Banerjee, Abhijit and Kaivan Munshi, "How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur," *The Review of Economic Studies*, 2004, *71* (1), 19–42.

Beaman, Lori, "649Social Networks and the Labor Market," in "The Oxford Handbook of the Economics of Networks," Oxford University Press, 04 2016.

_ and Jeremy Magruder, "Who gets the job referral? Evidence from a social networks experiment," *American Economic Review*, 2012, *102* (7), 3574–3593.

Bernard, Andrew B, Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes, "The origins of firm heterogeneity: A production network approach," *Journal of Political Economy*, 2022, *130* (7), 1765–1804.

Bertrand, Marianne and Antoinette Schoar, "The role of family in family firms," *Journal of economic perspectives*, 2006, *20* (2), 73–96.

Bhagavatula, Suresh, Manaswini Bhalla, Manisha Goel, and Balagopal Vissa, "Cultural homophily in corporate boards and firm," 2018.

Bharathi, Naveen, Deepak Malghan, Sumit Mishra, and Andaleeb Rahman, "Fractal urbanism: City size and residential segregation in India," *World Development*, 2021, *141*, 105397.

Bidner, Chris and Mukesh Eswaran, "A gender-based theory of the origin of the caste system of India," *Journal of Development Economics*, 2015, *114*, 142–158.

Boken, Johannes, Lucie Gadenne, Tushar Nandi, and Marta Santamaria, "Community networks and trade," *Warwick Economics Research Papers*, 2022.

Brinatti, Agostina and Nicolas Morales, "Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments," 2021.

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *The Quarterly journal of economics*, 2006, *121* (2), 541–585.

Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin, "Job search and hiring with limited information about workseekersâ™ skills," *American Economic Review*, 2022, *112* (11), 3547–3583.

Cassan, Guilhem, Daniel Keniston, and Tatjana Kleineberg, "A division of laborers: Identity and efficiency in India," Technical Report, National Bureau of Economic Research 2021.

Chandrasekhar, Arun G, Melanie Morten, and Alessandra Peter, "Network-Based Hiring: Local Benefits; Global Costs," Technical Report, National Bureau of Economic Research 2020.

Chaney, Thomas, "Distorted gravity: the intensive and extensive margins of international trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

__ , "The network structure of international trade," *American Economic Review*, 2014, *104* (11), 3600–3634.

Chaurey, Ritam, "Labor regulations and contract labor use: Evidence from Indian firms," *Journal of Development Economics*, 2015, *114*, 224–232.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer, "The trade-creating effects of business and social networks: evidence from France," *Journal of international Economics*, 2005, *66* (1), 1–29.

Deschênes, Olivier and Michael Greenstone, "The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather," *American economic review*, 2007, *97* (1), 354–385.

Deshpande, Manali S, "History of the Indian caste system and its impact on India today," 2010.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg, "The political economy of linguistic cleavages," *Journal of development Economics*, 2012, *97* (2), 322–338.

__ , Joseph F Gomes, and Diego Malo Rico, "Ethnic remoteness reduces the peace dividend from trade access," Technical Report, National Bureau of Economic Research 2023.

Dhillon, Amrita, Vegard Iversen, and Gaute Torsvik, "Employee referral, social proximity, and worker discipline: Theory and suggestive evidence from India," *Economic Development and Cultural Change*, 2021, *69* (3), 1003–1030.

Einav, Liran, Peter J Klenow, Jonathan D Levin, and Raviv Murciano-Goroff, "Customers and retail growth," Technical Report, National Bureau of Economic Research 2021.

Emerick, Kyle, "Trading frictions in Indian village economies," *Journal of Development Economics*, 2018, *132*, 32–56.

Felbermayr, Gabriel J and Farid Toubal, "Cultural proximity and trade," *European Economic Review*, 2010, *54* (2), 279–293.

Fernando, A Nilesh, Niharika Singh, and Gabriel Tourek, "Hiring frictions and the promise of online job portals: Evidence from india," *American Economic Review: Insights, forthcoming*, 2023.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig, "Cultural proximity and loan outcomes," *American Economic Review*, 2017, *107* (2), 457–492.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

__ , __ , and __ , "The slow growth of new plants: Learning about demand?," *Economica*, 2016, *83* (329), 91–129.

Fujiy, Brian Cevallos, Gaurav Khanna, and Hiroshi Toma, "Cultural Proximity and Production Networks," in "Research Seminar in International Economics, University of Michigan Working Papers" number 686 2022.

Ghosh, Arkadev, "Religious divisions and production technology: Experimental evidence from India," *Available at SSRN 4188354*, 2022.

Goraya, Sampreet Singh, "How Does Caste Affect Entrepreneurship? Birth versus Worth," *Journal of Monetary Economics*, 2023.

Gupta, Apoorv, "Demand for Quality, Variable Markups and Misallocation: Evidence from India," *Variable Markups and Misallocation: Evidence from India (January 1, 2020)*, 2020.

Hardy, Morgan and Gisella Kagy, "It's getting crowded in here: experimental evidence of demand constraints in the gender profit gap," *The Economic Journal*, 2020, *130* (631), 2272–2290.

Heath, Rachel, "Why do firms hire using referrals? Evidence from Bangladeshi garment factories," *Journal of Political Economy*, 2018, *126* (4), 1691–1746.

Helpman, Elhanan, Marc J Melitz, and Stephen R Yeaple, "Export versus FDI with heterogeneous firms," *American economic review*, 2004, *94* (1), 300–316.

Hjort, Jonas, "Ethnic divisions and production in firms," *The Quarterly Journal of Economics*, 2014, *129*

(4), 1899–1946.
— and Jonas Poulsen, "The arrival of fast internet and employment in Africa," *American Economic Review*, 2019, *109* (3), 1032–1079.
Hotelling, Harold, "Stability in competition," *The economic journal*, 1929, *39* (153), 41–57.
Hottman, Colin J, Stephen J Redding, and David E Weinstein, "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.
Hsieh, Chang-Tai and Peter J Klenow, "The life cycle of plants in India and Mexico," *The Quarterly Journal of Economics*, 2014, *129* (3), 1035–1084.
Hunt, Vivian, Sara Prince, Sundiatu Dixon-Fyle, and Lareina Yee, "Delivering through diversity," *McKinsey & Company*, 2018, *231*, 1–39.
Jayachandran, Seema, "Selling labor low: Wage responses to productivity shocks in developing countries," *Journal of political Economy*, 2006, *114* (3), 538–575.
Jensen, Robert and Nolan H Miller, "Market integration, demand, and the growth of firms: Evidence from a natural experiment in India," *American Economic Review*, 2018, *108* (12), 3583–3625.
Karlan, Dean, Markus Mobius, Tanya Rosenblat, and Adam Szeidl, "Trust and social collateral," *The Quarterly Journal of Economics*, 2009, *124* (3), 1307–1361.
Kaur, Supreet, "Nominal wage rigidity in village labor markets," *American Economic Review*, 2019, *109* (10), 3585–3616.
Madheswaran, Subramaniam and Paul Attewell, "Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey," *Economic and political Weekly*, 2007, pp. 4146–4153.
McCallum, John, "National borders matter: Canada-US regional trade patterns," *The American Economic Review*, 1995, *85* (3), 615–623.
Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 2003, *71* (6), 1695–1725.
Montalvo, Jose G and Marta Reynal-Querol, "Ethnic diversity and economic development," *Journal of Development economics*, 2005, *76* (2), 293–323.
Montalvo, José G and Marta Reynal-Querol, "Ethnic polarization, potential conflict, and civil wars," *American economic review*, 2005, *95* (3), 796–816.
Montgomery, James D, "Social networks and labor-market outcomes: Toward an economic analysis," *The American economic review*, 1991, *81* (5), 1408–1418.
Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended gravity," *The Review of economic studies*, 2019, *86* (6), 2668–2712.
MSME, *Fourth All India Census of Micro, Small and Medium Enterprises, 2006*, Ministry of Micro, Small and Medium Enterprises, 2009.
Munshi, Kaivan, "Networks in the modern economy: Mexican migrants in the US labor market," *The Quarterly Journal of Economics*, 2003, *118* (2), 549–599.
—, "Caste and the Indian economy," *Journal of Economic Literature*, 2019, *57* (4), 781–834.
— and Mark Rosenzweig, "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy," *American Economic Review*, 2006, *96* (4), 1225–1252.
— and —, "Networks and misallocation: Insurance, migration, and the rural-urban wage gap," *American Economic Review*, 2016, *106* (01), 46–98.
Nagavarapu, Sriniketh and Sheetal Sekhri, "Informal monitoring and enforcement mechanisms in public service delivery: Evidence from the public distribution system in India," *Journal of Development Economics*, 2016, *121*, 63–78.
OECD, *All hands in? Making diversity work for all*, OECD Publishing, 2020.
Oh, Suanna, "Does identity affect labor supply?," *American Economic Review*, 2023, *113* (8), 2055–2083.
Rauch, James E, "Business and social networks in international trade," *Journal of economic literature*, 2001, *39* (4), 1177–1203.
— and Vitor Trindade, "Ethnic Chinese networks in international trade," *Review of Economics and Statistics*, 2002, *84* (1), 116–130.
Santangelo, G., "Firms and Farms: The Local Effects of Farm Income on Firms' Demand," *Cambridge*

Working Papers in Economics, 2019.

Schmeiser, Katherine N, "Learning to export: Export growth and the destination decision of firms," *Journal of International Economics*, 2012, *87* (1), 89–97.

Spears, Dean and Amit Thorat, "The puzzle of open defecation in rural India: evidence from a novel measure of caste attitudes in a nationally representative survey," *Economic Development and Cultural Change*, 2019, *67* (4), 725–755.

Takeshita, Junko, Shiyu Wang, Alison W Loren, Nandita Mitra, Justine Shults, Daniel B Shin, and Deirdre L Sawinski, "Association of racial/ethnic and gender concordance between patients and physicians with patient experience ratings," *JAMA network open*, 2020, *3* (11), e2024583–e2024583.

Tan, Eugene and Teega H. Zeida, "Consumer Demand and Credit Supply as Barriers to Growth for Black-Owned Startups," *Accepted at Journal of Monetary Economics*, 2023.

A Model Derivations

A.1 Household Problem

The first-order conditions for the households are

$$\begin{aligned} \frac{a}{C_H - \overline{H}} - \lambda P_H &= 0 \\ (1-a)\frac{q(z(\omega),s,s^{'})c(z(\omega),s,s^{'})^{\frac{-1}{\sigma}}}{C_D^{\frac{\sigma-1}{\sigma}}} - \lambda p(z(\omega),s,s^{'}) &= 0 \\ P_H C_H + \sum_{s'} p(z(\omega),s,s^{'})c(z(\omega),s,s^{'}) &= I_s \end{aligned} \tag{A.1}$$

where λ is the Lagrange multiplier for the constraint on the total expenditure as defined in Equation (4). This specification implies that the representative household first allocates $P_H\overline{H}$ amounts of income to $\overline{H}$ units of homogeneous good, and then allocates the remaining income to the two goods proportional to their weights in the utility function. Manipulating the Equation (A.1), we get that

$$\begin{aligned} P_H C_H &= P_H\overline{H} + a(I_s - P_H\overline{H}) \\ \sum_j p(z(\omega),s,s^{'})c(z(\omega),s,s^{'}) &= (1-a)(I_s - P_H\overline{H}) \\ \lambda &= \frac{1}{I_s - P_H\overline{H}} \end{aligned} \tag{A.2}$$

Deriving the demand for a variety with quality $z(\omega)$, we get an iso-elastic residual demand curve

$$c(z(\omega),s,s^{'}) = y(z(\omega),s,s^{'}) = q(z(\omega),s,s^{'})^{\sigma}p(z(\omega),s,s^{'})^{-\sigma}C_D P_D^{\sigma}. \tag{A.3}$$

A.2 Firm Problem

A.2.1 Proof Proposition 1: Optimal Location

Firm-level profits of the firm owner of caste s' are given by

$$\Pi_D(z,s,\Delta\mathcal{X}_s) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_{s'}\right)^{1-\sigma} z^{\sigma}\sum_{s'}\Psi(d_{ss'},\Delta d_{s'})^{\sigma}\kappa_{s'} - \Phi(\Delta\mathcal{X}_{s'};\Gamma_{s'}) - \mathcal{F}^d_{s'}. \tag{A.4}$$

This combined with the expression $\Psi(d_{ss'}, \Delta d_{s'})^{\sigma} = 1 - \hat{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right)$, with $\beta = \hat{\beta}\sigma$ and $\Phi(\Delta\mathcal{X}_{s'};\Gamma_{s'}) = \sum_{k\in\mathcal{N}}\gamma_{s',k}\Delta\mathcal{X}^2_{s',k}$, we get

$$\begin{aligned}\Pi_D(z, s', \Delta\mathcal{X}_{s'}) = &\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_{s'}\right)^{1-\sigma} z^{\sigma}\sum_{s}\left(1 - \beta\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right)\right)\kappa_s \\ &- \sum_{k\in\mathcal{N}}\gamma_{s',k}\Delta\mathcal{X}^2_{s',k} - \mathcal{F}^d_{s'}.\end{aligned} \tag{A.5}$$

Using the fact $d_{ss,k} = 0$ and define $B_s = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma}$, we can write

$$\begin{aligned}\Pi_D(z, s', \Delta\mathcal{X}_s) = &B_{s'}\sum_s \kappa_s - B_{s'}\beta\sum_s\kappa_s\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right) \\ &- \sum_{k\in\mathcal{N}}\gamma_{s',k}(d_{s's',k} - \Delta\mathcal{X}_{s',k})^2 - \mathcal{F}^d_{s'}.\end{aligned} \tag{A.6}$$

Collecting terms and a few steps of Algebra give us

$$\Pi_D(z, s', \Delta\mathcal{X}_s) = B_{s'}\Upsilon\left(1 - \sum_s\sum_k \lambda_{ss',k}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right) - \mathcal{F}^m_{s'} - \mathcal{F}^d_{s'}, \tag{A.7}$$

where $B_{s'} = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_{s'}\right)^{1-\sigma} z^{\sigma}$ and $\Upsilon = \sum_{s\in\mathcal{S}}\kappa_s$, and $\lambda_{ss',k} = \frac{\beta\kappa_s + B^{-1}_{s'}\gamma_{s',k}\mathbb{1}_{s=s'}}{\Upsilon}$. $\mathbb{1}_{s=s'}$ is the indicator function that have value one if $s = s'$. Under these definitions, we can rewrite the profit maximisation as reduced down to an optimal (minisum Fermat-Weber Problem) location problem.[31]

$$\mathcal{V}_D(\Delta\mathcal{X}_{s',1}, ..., \Delta\mathcal{X}_{s',\mathcal{N}}) = \min_{\Delta\mathcal{X}_{s'}}\sum_s\sum_k \lambda_{ss',k}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2 \tag{A.8}$$

First-order conditions are given by

$$-\frac{\partial}{\partial\Delta\mathcal{X}_{s',k}}\sum_{s'}\lambda_{ss',k}(d_{ss',k} - \Delta\mathcal{X}_{s',k}) = 0 \quad \forall \quad k \in \mathcal{N} \tag{A.9}$$

This gives us the expression for the optimal distance moved

$$\Delta\mathcal{X}_{s',k} = \frac{\sum_{s\in\mathcal{S}}\lambda_{ss'k}d_{ss',k}}{\sum_{s\in\mathcal{S}}\lambda_{ss'k}}, \quad \forall \quad k \in \mathcal{N}. \tag{A.10}$$

[31] This is a version of Fermat-Weber Problem as our problem is quadratic in distance. This allows us to have the closed-form solution to the opimal location.

A.2.2 Taste Shifters and Transportation Costs.

The revenues are given by

$$r(z,s,s') = \tau_{ss'}^{1-\sigma} r(z,s,s), \quad \pi(z,s,s') = \frac{r(z,s,s')}{\sigma}. \tag{A.11}$$

Total firm-level profits can be written as

$$R(z,s') = \sum_s r(z,s,s^{'}) = \left(\frac{\sigma}{\sigma-1}\mathcal{C}_{s'}\right)^{1-\sigma} z^{\sigma} \sum_s \tau_{ss'}^{1-\sigma} \Psi(d_{ss'})^{\sigma} \kappa_s \tag{A.12}$$

$$\Pi_D(z,s') = \sum_s \pi(z,s,s^{'}) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_{s'}\right)^{1-\sigma} z^{\sigma} \sum_s \tau_{ss'}^{1-\sigma} \Psi(d_{ss'})^{\sigma} \kappa_s - \mathcal{F}_{s'}^{d} \tag{A.13}$$

We assume, when firms change their ethnic identity, that also translates into a change in the physical distance, they are perfectly correlated. This assumption makes the model tractable.

We assume $\Psi(d_{ss'}, \Delta d_{s'}) = e^{-\hat{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s',k})^2\right)}$ and $\tau(d_{ss'}, \Delta d_{s'}) = e^{\gamma\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s',k})^2\right)}$ and define $\tilde{\beta} = \hat{\beta}\sigma + \gamma(\sigma-1)$. We assume the first-order Taylor approximation of

$$\tau(d_{ss'}, \Delta d_{s'})^{1-\sigma}\Psi(d_{ss'}, \Delta d_{s'})^{\sigma} = 1 - \tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s',k})^2\right). \tag{A.14}$$

The solution to the optimal location problem is similar to Equation A.10, with the only difference being that the $\tilde{\beta}$, which captures the rate at which trade between two castes declines with distance is a combination of taste parameter β and transportation cost parameter γ.

A.2.3 A side note on Pareto

$$H(x) := \int_x^{\infty} a^{\sigma} dG(a)$$

$$\begin{aligned}
H(x) &= \int_x^{\infty} a^{\sigma} \xi a^{-(\xi+1)} da \\
&= -\left.\frac{\xi}{\xi-(\sigma)} a^{-(\xi-(\sigma))}\right|_x^{\infty} \\
&= \frac{\xi}{\xi-(\sigma)} x^{-(\xi-(\sigma))}, \quad \text{assuming } \xi > \sigma
\end{aligned}$$

A.2.4 A side note on Price index

Using the downward sloping demand curve derived in the consumer problem in Equation A.3

$$\begin{aligned}
c(\omega, s, s') &= p(\omega, s, s')^{-\sigma} q(\omega, s, s')^{\sigma} \kappa_s \\
p(\omega, s, s') c(\omega, s, s') &= p(\omega, s, s')^{1-\sigma} q(\omega, s, s')^{\sigma} \kappa_s \\
P_{D,s} C_{D,s} \equiv \int p(\omega, s, s') c(\omega, s, s') d\omega &= \int_{\Omega_s} p(\omega, s, s')^{1-\sigma} q(\omega, s, s')^{\sigma} P_{D,s}^{\sigma} C_{D,s} d\omega \\
P_{D,s} C_{D,s} &= P_{D,s}^{\sigma} C_{D,s} \int_{\Omega_s} p(\omega, s, s')^{1-\sigma} q(\omega, s, s')^{\sigma} d\omega \\
P_{D,s} &= \left(\int_{\Omega_s} q(\omega, s, s')^{\sigma} p(\omega, s, s')^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}
\end{aligned} \tag{A.15}$$

A.2.5 Derivation of Firm-level Revenues

Using $\kappa_s = C_{D,s} P_{D,s}^{\sigma} = C_{D,s} P_{D,s} P_{D,s}^{\sigma-1} = (1-a)(I_s - \overline{H}) P_{D,s}^{\sigma-1}$, define $\hat{I}_s = (1-a)(I_s - \overline{H})$. Now, we can rewrite the firm-level revenues from a consumer of caste s with $\kappa_s = \hat{I}_s P_{D,s}^{\sigma-1}$ and revenues are

$$r(z, s, s^{'}) = q(z, s, s^{'})^{\sigma} \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_s' \tau_{ss'} \right)^{1-\sigma} \kappa_s \tag{A.16}$$

Now, we can compute the cross-caste micro trade elasticities. Taking logs on both sides and taking derivative w.r.t $\log I_s$, and $\Lambda(d^*_{ss'}) = e^{-\tilde{\beta} \sum_{k \in \mathcal{K}} (d^*_{ss',k})^2}$, we have

$$\frac{\partial \log r(z, s, s^{'})}{\partial \log I_s} = \underbrace{\frac{\partial \log \kappa_s}{\partial \log I_s}}_{\textbf{Size effect}} + \underbrace{2\tilde{\beta} \sum_{k \in \mathcal{K}} d^*_{ss',k} \frac{\partial \Delta \mathcal{X}_{s',k}}{\partial \log I_s}}_{\Delta \textbf{ Optimal Distance}} \tag{A.17}$$

A.2.6 Proof of Proposition 2

Defining overall firm-level resistent term as $\Lambda(d_{ss'}, \Delta d) = \tau_{ss'}^{1-\sigma} \Psi(d_{ss'}, \Delta d)$

$$R(z, s') = \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_{s'} \right)^{1-\sigma} z^{\sigma} \sum_s \kappa_s \Lambda(d_{ss'}, \Delta d) \tag{A.18}$$

$$R(z, s') = \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_{s'} \right)^{1-\sigma} z^{\sigma} \sum_s \kappa_s e^{-\tilde{\beta} \left(\sum_{k \in \mathcal{N}} (d_{ss',k} - \Delta \mathcal{X}_{s,k})^2 \right)} \tag{A.19}$$

$$\log R(z, s') = \alpha_{s'} + \log \sum_s \kappa_s e^{-\tilde{\beta} \left(\sum_{k \in \mathcal{N}} (d_{ss',k} - \Delta \mathcal{X}_{s,k})^2 \right)}, \tag{A.20}$$

where $\alpha_{s'} = \log\left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma}$. Define the share of sales of the firm owner of caste s' to any caste s as

$$A_{ss'} = \frac{\left(\kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s,k})^2\right)}\right.}{\sum_s \kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s,k})^2\right)}} \tag{A.21}$$

Taking the derivative both of the revenue equation, we get

$$\frac{\partial \log R(z,s')}{\partial \log I_s} = \frac{\partial \log \kappa_s}{\partial \log I_s} A_{ss'} + 2\tilde{\beta}\sum_{k\in\mathcal{K}} d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial \log I_s} A_{ss'}, \tag{A.22}$$

$$\frac{\partial \log R(z,s')}{\partial \log I_s} = \frac{\partial \log r(z,s,s')}{\partial \log I_s} A_{ss'}. \tag{A.23}$$

We also know $\Delta\mathcal{X}_{s,k} = \frac{\sum_{s'\in\mathcal{S}}\tilde{\lambda}_{ss'k}d_{ss',k}}{\sum_{s'\in\mathcal{S}}\tilde{\lambda}_{ss'k}}$, and we have $\tilde{\lambda}_{ss',k} = \frac{\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s',k}\mathbb{1}_{s=s'}}{\mathrm{Y}}$ and $\mathrm{Y} = \sum_{s\in\mathcal{S}}\kappa_s$. We can show that

$$\begin{aligned}
\frac{\partial}{\partial \log I_s}\frac{\sum_{s\in\mathcal{S}}\tilde{\lambda}_{ss',k}d_{ss',k}}{\sum_{s\in\mathcal{S}}\tilde{\lambda}_{ss',k}} =& \frac{\partial}{\partial \log I_s}\frac{\sum_{s\in\mathcal{S}}(\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'})d_{ss',k}}{\sum_{s\in\mathcal{S}}(\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'})} \\
=& \frac{\tilde{\beta}\frac{\partial\kappa_s}{\partial\log I_s}d_{ss',k}}{\sum_{s\in\mathcal{S}}(\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s',k}\mathbb{1}_{s=s'})} - \frac{\tilde{\beta}\frac{\partial\kappa_s}{\partial\log I_s}\sum_{s'\in\mathcal{S}}(\tilde{\beta}\kappa_{s'} + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'})d_{ss',k}}{\left(\sum_{s'\in\mathcal{S}}(\tilde{\beta}\kappa_{s'} + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'})\right)^2} \\
=& \frac{\tilde{\beta}}{\mathrm{Y}\sum_{s\in\mathcal{S}}\tilde{\lambda}_{ss',k}}\frac{\partial\kappa_s}{\partial\log I_s}\left(d_{ss',k} - \frac{\sum_{s\in\mathcal{S}}\tilde{\lambda}_{ss',k}d_{ss',k}}{\sum_{s\in\mathcal{S}}\tilde{\lambda}_{ss',k}}\right) \\
=& \frac{\tilde{\beta}}{\mathrm{Y}\sum_{s\in\mathcal{S}}\tilde{\lambda}_{ss',k}}\frac{\partial\kappa_s}{\partial\log I_s}d^*_{ss',k}
\end{aligned} \tag{A.24}$$

Under fully exogenous cultural proximity with $d_{ss} = 0$ and $d_{ss'} > 0$, the revenue elasticity to demand shocks is given by

$$\frac{\partial \log R(z,s')}{\partial \log I_s} = \frac{\partial \log \kappa_s}{\partial \log I_s} A_{ss'} \quad \forall \quad s \neq s'. \tag{A.25}$$

A.2.7 Proof of Proposition 3

We have computed the revenue elasticity. Here we show under what conditions the revenue elasticity of firm owner s is increasing in the income shock to the consumers of the same caste. We first formulate the general problem. The revenue shares of the firm owner of caste s' are defined as

$$A_{ss'} = \frac{\kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}}{\sum_s \kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}} \tag{A.26}$$

The derivative of revenue shares to the overall resistance parameter $\tilde{\beta}$ is

$$\begin{aligned}\frac{\partial A_{ss'}}{\partial \tilde{\beta}} =& A_{ss'}\left(-\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right) + 2\tilde{\beta}A_{ss'}\sum_{k\in\mathcal{K}} d^*_{ss',k}\frac{\partial \Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}} \\ &+ A_{ss'}\frac{\sum_s \kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2)}{\sum_s \kappa_s e^{-\tilde{\beta}\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2}} \\ &- A_{ss'}\frac{\sum_s \kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}2\tilde{\beta}\sum_{k\in\mathcal{K}} d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}}}{\sum_s \kappa_s e^{-\tilde{\beta}\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2}}\end{aligned} \tag{A.27}$$

We can collect 1 and 3 terms and 2 and 4 terms and rewrite this as

$$\frac{\partial A_{ss'}}{\partial \tilde{\beta}} = -A_{ss'}\,(d^*_{ss'})^2 + A_{ss'}\sum_s A_{ss'}(d^*_{ss'})^2 + 2\tilde{\beta}A_{ss'}\sum_{k\in\mathcal{K}} d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}} - 2\tilde{\beta}A_{ss'}\sum_s A_{ss'}\sum_{k\in\mathcal{K}} d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}} \tag{A.28}$$

In the case of exogenous caste distance, $\frac{\partial A_{ss'}}{\partial\tilde{\beta}} > 0$ as $d_{ss} = 0$, and last two terms are zero as well. In the endogenous case, $d_{ss} \neq 0$, we need more assumptions. First, if we assume, the costs are proportional to $\tilde{\beta}$ such that $\gamma_{s',k} = \tilde{\beta}\tilde{\gamma}_{s',k}$, then $\lambda_{ss',k}$ and thus $\Delta\mathcal{X}_{s',k}$ are independent of $\tilde{\beta}$. Therefore, the last two terms of the previous equation are zero. In this case

$$\frac{\partial A_{ss'}}{\partial \tilde{\beta}} = -A_{ss'}\,(d^*_{ss'})^2 + A_{ss'}\sum_s A_{ss'}(d^*_{ss'})^2 \tag{A.29}$$

Using the fact that $\sum_s A_{ss'} = 1$, we can rewrite this as

$$\frac{\partial A_{ss'}}{\partial \tilde{\beta}} = A_{ss'}\sum_{s''} A_{s''s'}\left((d^*_{s''s'})^2 - (d^*_{ss'})^2\right) \tag{A.30}$$

A sufficient conditions for A_{ss} to be increasing in $\tilde{\beta}$ is that the $d^*_{ss} \leq d^*_{ss'}$ for $s \neq s'$. In other words, the cost of adjusting distance should be high enough.

Finally, if the share of revenues coming from the firm's own caste consumer is increasing in $\tilde{\beta}$, then the firm is selling less to other castes, $1 - A_{ss}$, therefore, the revenue elasticity to the income shock is lower as well.

B Empirical Analysis

B.1 Data

B.1.1 MSME

Our main source of data is the Micro, Small and Medium Enterprises (MSME) data (MSME, 2009).

Main variables. The MSME data set is based on MSME sector which is defined by the Micro, Small and Medium Enterprise Development (MSMED) act of 2006, spans the non-agricultural enterprises of the economy and contains a representative sample of the *MSMEs* that invest less than INR 100 million (manufacturing sector) or INR 50 million (services sector).

In particular, the act notified the following enterprises, whether proprietorship, Hindu undivided family, an association of persons, co-operative society, partnership or undertaking or any other legal entity, by whatever name called:- In case of enterprises engaged in manufacturing or production of goods pertaining to any industry specified in the First Schedule to the Industries (Development and Regulation) Act, 1951, as: (i) a micro enterprise, where the investment in plant and machinery does not exceed 2.5 million rupees, (ii) a small enterprise, where the investment in plant and machinery is more than 2.5 million but does not exceed 50 million rupees; or (iii) a medium enterprise, where the investment in plant and machinery is more than 50 million rupees but does not exceed 100 million rupees. In the case of the enterprises engaged in providing or rendering of services, as: (i) a micro enterprise, where the investment in equipment does not exceed 1 million rupees; (ii) a small enterprise, where the investment in equipment is more than 1 million rupees but does not exceed 20 million rupees; or (iii) a medium enterprise, where the investment in equipment is more than 20 million rupees but does not exceed 50 million rupees.

According to the 4th MSME data set of India 2006, the MSME sector accounts for 37% of the manufacturing output and 89% of the total employment in the manufacturing sector. The sector is estimated to employ about 59 million individuals in over 26.1 million units throughout the country. Further, 1.5 million (5.94%) are registered MSMEs and 24.5 million (94.06 %) are unregistered MSMEs that employ 16.62 % and 83.38 % of the workforce respectively.

The MSME data set has two parts: a census of registered MSMEs and a sample survey of unregistered MSMEs. A total number of 126,169 enterprises are surveyed to capture a representative sample of unregistered MSMEs. There are 1.65 million observations in total, of which 179,041 do not provide information on the wage bill. Therefore, 1.45 million observations (1.12 million in the manufacturing sector) are left after the cleaning process. HC, MC, and LC represent 51%, 39.1%, and 9.8% of observations in the sample. In terms of total output, HC, MC, and LC enterprises produce 61.3%, 30.7% and 8%, respectively, whereas in terms of total credit, 75.1% is allocated to HC enterprises and only 4.7% to LC enterprises. The enterprise-level statistics are provided in Panel A of Table B.1.

Table B.1: Summary statistics

Panel A: MSME 2006-07 - Firm-level statistics								
	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Employment	3.14	(19.04)	2.32	(6.20)	1.98	(7.55)	2.59	(12.98)
Revenues (in thousands)	1,144	(35000)	263	(12200)	262	(19900)	608	(24600)
Materials (in thousands)	697	(23000)	128	(6079)	135	(13000)	352	(15800)
Loans (in thousands)	429	(6427)	94	(236)	59	(100)	209	(3869)

Panel B: NSS 2004-2010 - Individual-level statistics								
	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Wages	148.42	(295.11)	79.23	(105.71)	59.47	(78.08)	89.15	(171.70)
Sex (male)	0.81	(0.39)	0.73	(0.44)	0.70	(0.46)	0.74	(0.44)
Age	35.36	(12.01)	35.01	(12.42)	34.67	(12.56)	34.97	(12.37)
Education	6.92	(3.90)	5.04	(3.60)	4.01	(3.29)	5.13	(3.74)
Land owned	143.32	(6326.33)	126.90	(591.93)	96.29	(407.48)	119.25	(3151.83)
Employed in agri.	0.25	(0.43)	0.45	(0.50)	0.57	(0.49)	0.45	(0.50)

Panel C: NSS 2004-2008 - Household-level statistics								
	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
MPCE	1234.4	(1288.3)	807.2	(822.8)	627.1	(465.0)	883.6	(947.4)
Education	7.49	(4.08)	5.54	(3.89)	4.54	(3.73)	5.84	(4.07)
Meals	1.33	(0.49)	1.42	(0.52)	1.37	(0.51)	1.38	(0.51)

Notes: The table reports descriptive statistics. LC, MC, and HC represent entrepreneurs historically classified as belonging to the historically disadvantaged castes, middle-castes, and historically privileged castes, respectively. Sampling multipliers are applied in all the panels.
Panel A reports statistics for the MSME data set. Each row reports summary statistics for HC, MC, LC, and the full sample. Total refers to the total number, Mean refers to the mean value, and S.D. refers to the standard deviation. Employment is measured as the number of employees (e.g., the average number of employees for HC enterprises is 3.14), and revenue, materials, and loans are in Indian rupees. Output is gross sales, labour is total wages, and Materials is material input value.
Panel B reports statistics for the NSS Employment and Unemployment data set. S.D. is the standard deviation. Each row reports summary statistics for HC, MC, LC, and the full sample. Mean refers to the mean value. Wages is measured as the daily wage of workers, deflated for inflation, in Indian rupees. Education takes whole numbers between 0 (not literate) and 14 (post-graduate and above).
Panel C reports household-level statistics for the NSS Household Consumption Expenditure data set. S.D. is the standard deviation. Mean refers to the mean value. The variable *MPCE* reports the monthly per-capita expenditure, deflated for inflation, in Indian rupees. Education refers to the education of the head of the household and is reported in whole numbers between 0 (not literate) and 14 (post-graduate and above). Meals refers to the number of meals eaten by the head of the household in a day.

Additional variables: Panel A in Table B.3 shows the average prices charged and quantities sold for the primary product sold by the firm and the primary raw material used by the firm. It shows that the average product price and input material price is higher for HC firms than for LC firms. Figure B.2 plots the full distribution of product prices (across and within district and product category) and shows that the HC and LC firms exhibit a similar distribution, with LC firms's curve shifted to the left. Motivated by the difference in prices, we parameterise the notion of product quality in our quantitative model. HC firms also show larger input material procurement and sales than LC firms, consistent with the previous observation that HC firms operate at a larger size than LC firms.

Panel B in Table B.3 shows the caste-wise employment for firms belonging to different castes. We see a very distinct positive assortative employer-employee matching across castes. This further enhances the notion of ethnic identity of firms, which we parameterise in our quantitative model.

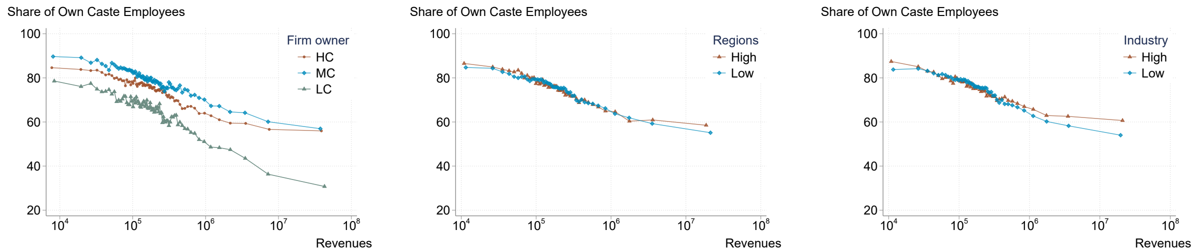
The MSME datset also provides retrospective information on sales and input materials for the enterprises that survive up to 2006-07. This allows us to construct a balanced panel of MSMEs for the three-year period, 2004 - 2007. This panel information allows us to compute statistics such as auto-correlation in output and materials. Panel C in Table B.3 shows that the auto-correlation in output is much higher for LC firms as they struggle to grow, while the auto-correlation in input materials is similar across castes.

Table B.2: Size Distribution and Cross-Caste Hiring

All firms	Mean	Median	p5	p95	N
Employees	5.5	2	1	15	595973
EMP LC	1.1	0	0	4	595973
EMP MC	2.0	1	0	5	595973
EMP HC	2.4	1	0	6	595973
HC firms					
Employees	7.6	2	1	20	259520
EMP LC	1.4	0	0	5	259520
EMP MC	1.8	0	0	6	259520
EMP HC	4.3	2	0	10	259520
MC firms					
Employees	4.0	2	1	10	258584
EMP LC	0.6	0	0	2	258584
EMP MC	2.6	2	0	6	258584
EMP HC	0.8	0	0	3	258584
LC firms					
Employees	3.7	2	1	8	77869
EMP LC	1.8	1	0	4	77869
EMP MC	0.7	0	0	2	77869
EMP HC	1.1	0	0	3	77869

Notes: The table reports descriptive statistics. LC, MC, and HC represent entrepreneurs historically classified as belonging to the historically disadvantaged castes, middle-castes, and historically privileged castes, respectively. Sampling multipliers are applied in all the panels.

Figure B.1: Size Distribution and Cross-Caste Hiring



(a) By Caste (b) By Region (c) By Sector

Notes: The table reports descriptive statistics. LC, MC, and HC represent entrepreneurs historically classified as belonging to the historically disadvantaged castes, middle-castes, and historically privileged castes, respectively. Sampling multipliers are applied in all the panels.

B.1.2 Rainfall

The Tropical Rainfall Measuring Mission (TRMM) provides gridded rainfall rates at very high spatial and temporal resolution. Daily rainfall measures are available at the 0.25 by 0.25 degree grid-cell size. Spatially, we aggregate this data by calculating the mean rainfall registered on the grid points within the boundary of a district. Temporally, we aggregate this data as the total rainfall during the months of June, July, August and September. Together, we obtain a measure of rainfall at the district-year level.[32] For Indian districts between 1980 and 2016, monsoon rainfall accounts for more than 75% of annual rainfall[33]. Figure 3b plots the spatial variation in rainfall, measured in millimeters of monsoon rainfall.

[32] See also, for instance, Deschênes and Greenstone (2007) and Santangelo (2019) for previous usage of seasonal rainfall data as the measure of precipitaion at the county (in USA) or district (in India) level.

[33] This remains a robust feature for different specifications of rainfall. For example, monsoon rainfall share in total yearly rainfall is more than 70% in the specification used by Santangelo (2019).

Table B.3: MSME: Additional summary statistics

Panel A: MSME 2004 - 2007 - Firm-level price and quantity statistics

	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
primary product price	4323	(410718)	2915	(3273214)	1544	(381362)	3256	(2228051)
primary input material price	20691	(2645055)	9100	(858262)	11729	(2588369)	14629	(2104261)
primary product quantity	14928	(734931)	4839	(95691)	8245	(155886)	9312	(468435)
primary input material quantity	5509	(143512)	1581	(56139)	1669	(57579)	3132	(10058)

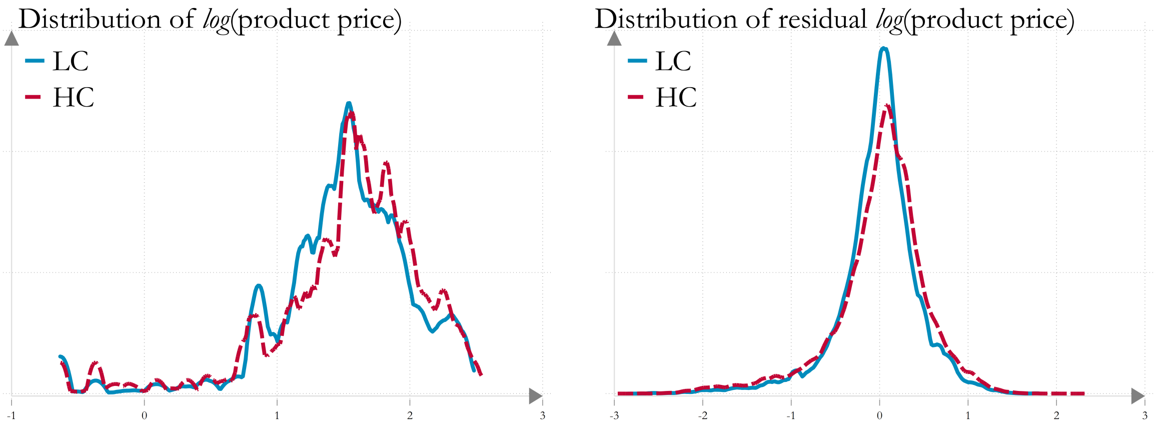
Panel B: MSME 2004 - 2007 - Firm-level employment statistics

	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
HC workers	**2.45**	(14.3)	0.98	(4.9)	0.83	(7.98)	1.95	(12.2)
MC workers	1.39	(9.9)	**1.96**	(3.6)	0.75	(6.4)	1.71	(6.0)
LC workers	0.98	(11.4)	0.44	(6.0)	**1.34**	(4.6)	0.88	(8.7)

Panel C: MSME 2004 - 2007 - Firm-level panel statistics

	HC		MC		LC		All	
		(S.E.)		(S.E.)		(S.E.)	Mean	(S.E.)
Auto-correlation in								
output	0.39	(0.12)	0.39	(0.07)	0.47	(0.10)	0.39	(0.10)
input materials	0.38	(0.07)	0.37	(0.07)	0.38	(0.07)	0.38	(0.06)

Notes: The table reports descriptive statistics. LC, MC, and HC represent entrepreneurs historically classified as belonging to the historically disadvantaged castes, middle-castes, and historically privileged castes, respectively. Sampling multipliers are applied in all the panels.
Panel A reports statistics for the MSME data set. Each row reports summary statistics for HC, MC, LC, and the full sample. Primary product is the main product produced by the firm as reported in the survey, and primary input material is the main input material used by the firm as reported in the survey. Mean refers to the mean value and S.D. refers to the standard deviation.
Panel B reports statistics for the MSME data set. Each row reports summary statistics for HC, MC, LC, and the full sample. Workers refer to the number of employees. Mean refers to the mean value and S.D. refers to the standard deviation.
Panel A reports statistics for the MSME data set. Each row reports summary statistics for HC, MC, LC, and the full sample. S.E. refers to the standard error. S.E. are clustered at the district-level. Output is measured using gross sales and input materials is measured using the total value of inputs.



(a) unconditional

(b) within District and Product-Category

Figure B.2: Distribution of product prices: MSME 2006-07

Notes: This figure uses MSME data from 2006-07 to plot the distribution of firm-level product prices. Panel (a) plots the unconditional distribution, and Panel (b) plots the residual distribution of prices, after controlling for district and year fixed-effects.

B.2 Geographical segregation

The premise of our study is to investigate whether firms belonging to a certain ethnic group tend to cater to consumers of the same group, and hence suffer from the consequences to growth due to limited demand. In Section 5, we have documented the presence of such ethnic linkages and the resulting segregation in the product market. This overall segregation may be a combination of two underlying channels: (a) homophilic preferences, where consumers have a preference to buy from firms belonging to ethnically similar groups, and (b) geographical distance, where consumers have a cost to access products that geographically farther.[34]

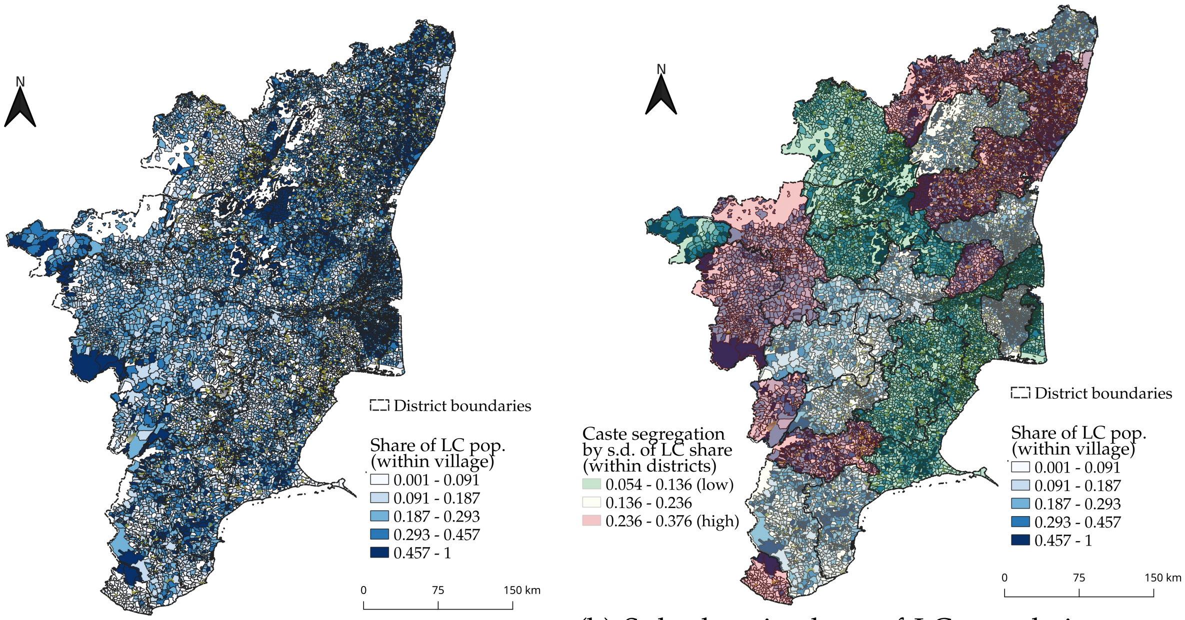
To understand whether our results in Section 5 are driven by the historical geographical segregation of ethnic groups within districts, we carry out the following exercise. We collect population data for different castes at the village level from the Â°UG database (Asher and Novosad, 2019). Within each village, we obtain the share of LC (=SC+ST) population. To measure geographical segregation across castes within district, we use the standard deviation in the share of LC population in villages within a district:

$$Segregation_d = sd(LCshare_v) \tag{B.31}$$

where d denotes district and v denotes village and sd denotes the standard deviation function.

[34]Recent papers (see, for instance Jensen and Miller, 2018 and Asturias, García-Santana and Ramos, 2019) have shown that geographical distance can be an important factor in determining the market of a firm's product.

Figure B.3: Geographical segregation, by caste, across districts of Tamil Nadu (2001)



(a) Share of LC population across TN villages

(b) Std. dev. in share of LC population across villages within TN districts

Notes: This figure uses Socioeconomic High-resolution Rural-Urban Geographic Data Platform for India (SHRUG) data's Economic and Population Census module (2001) to plot the standard deviation in the share of LC (SC+ST) population across villages within each district across a state in India, that is Tamil Nadu. Panel (a) plots the share of LC population in each village in Tamil Nadu and Panel (b) overlays on top of the district-level standard deviation in LC population share across villages within each district.

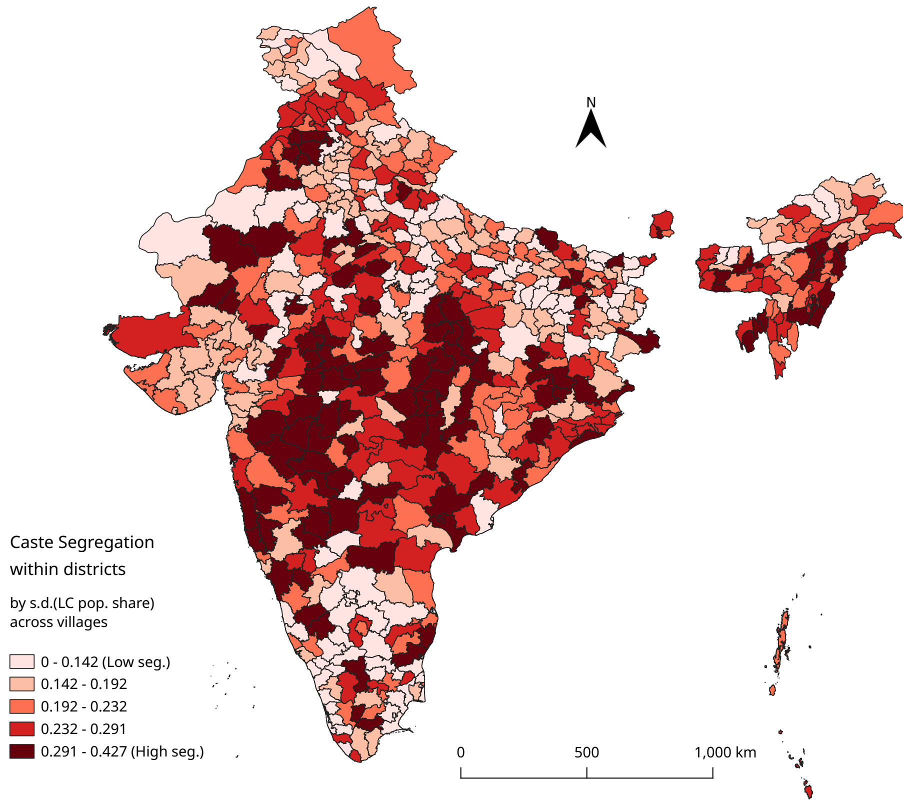


Figure B.4: Geographical segregation, by caste, across districts of India (2001)

Notes: This figure uses Socioeconomic High-resolution Rural-Urban Geographic Data Platform for India (SHRUG) data's Economic and Population Census module (2001) to plot the standard deviation in the share of LC (SC+ST) population across villages within each district across India.

Table B.4: Elasticities in *unsegregated* rural India

	Individual wages			Household consumption				Firm outcomes	
	(1) All	(2) Agri.	(3) Non-Agri.	(4) All	(5) Mfg.	(6) Services	(7) Durables	(8) Revenue	(9) Material Input
log(rainfall)	0.078***	0.062***	0.077***	0.005	-0.038	0.026	-0.097	-	-
	(0.023)	(0.023)	(0.022)	(0.026)	(0.031)	(0.056)	(0.064)	-	-
MC	-0.009	-0.204	0.159	-0.607***	-0.568***	-1.336***	-1.152**	-1.974***	-2.380***
	(0.160)	(0.140)	(0.156)	(0.173)	(0.132)	(0.437)	(0.520)	(0.350)	(0.438)
LC	-0.149	-0.325***	-0.144	-1.010***	-0.749***	-1.667***	-2.581***	-2.229***	-2.658***
	(0.133)	(0.117)	(0.144)	(0.161)	(0.184)	(0.412)	(0.476)	(0.387)	(0.505)
log(rainfall) × MC	-0.018	0.027	-0.040*	0.059**	0.060***	0.145**	0.139*	0.235***	0.283***
	(0.024)	(0.021)	(0.023)	(0.026)	(0.020)	(0.065)	(0.078)	(0.050)	(0.063)
log(rainfall) × LC	-0.001	0.033**	-0.015	0.100***	0.048**	0.138***	0.280***	0.241***	0.280***
	(0.017)	(0.016)	(0.020)	(0.021)	(0.024)	(0.051)	(0.056)	(0.057)	(0.074)
Observations	87,211	42,660	44,551	112,822	112,346	97,232	91,387	897,840	890,182
R-squared	0.408	0.365	0.386	0.394	0.263	0.246	0.222	0.431	0.467
Controls$_{it}$	✓	✓	✓	-	-	-	-	-	-
District FE	✓	✓	✓	✓	✓	✓	✓	-	-
Year FE	✓	✓	✓	✓	✓	✓	✓	-	-
District × Year FE	-	-	-	-	-	-	-	✓	✓
Product FE	-	-	-	-	-	-	-	✓	✓

Notes. In Columns (1) to (3), the regressions are of individual-level daily wage (logarithmic) on rainfall (logarithmic). $Agri.$ stands for agricultural workers in rural areas. $Non - Agri.$ stands for workers employed in sectors non-agricultural sectors such as manufacturing, constructions, services, etc. in rural areas. The additional individual-level control variables included are age, gender, education, land possessed, and crop season. Sample includes individuals between the ages of 18 and 60. In Columns (4) to (7), the regressions are of households' monthly per capita consumption (logarithmic) on rainfall (logarithmic). $Mfg.$ stands for consumption of clothing and footwear. In Columns (8) and (9), the regressions are of firm-level variables (logarithmic) on rainfall (logarithmic). Sample omits observations of districts with standard deviation in village-level LC population share in the top quartile. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at district level in all regressions, *** p<0.01, ** p<0.05, * p<0.1.

Figure B.4 shows that there is a lot of variation in within-district geographical segregation across India. Figure B.3 provides a more granular look at this measure by focusing on the state of Tamil Nadu. Having constructed this measure, we use its spatial variation and replicate our firm-level empirical analysis. We omit observations of highly *geographically segregated* districts, that is those districts that belong to the top quartile of our measure of segregation. Table B.4 shows that the results are in line with our baseline results from Table 3.

This evidence suggests that geographical segregation may not be the driving factor behind the homophilic demand patterns observed in Table 3. However, note that the interaction coefficients for LC firms in Columns (8) and (9) are larger than the baseline, suggesting that in districts with low geographical frictions, stronger homophilic patters can lead to larger revenues for LC firms. Therefore, we incorporate the presence of geographical frictions to obtain a more complete picture of the Indian rural economy and parameterise the forces due to it in our quantitative model (see Table **??** for more details).

B.3 Additional evidence

In this subsection, we further investigate the nature of growth shown by LC firms and the mechanism through which higher rainfall relaxes the constraints of an LC firm.

Could the results on firm growth be explained by a relaxation of other variable inputs? Could the results on firm growth be explained by reasons other than relaxation of working-capital constraints? To answer this, we present for additional evidence on the mechanism. We implement a cross-sectional regression, using the 2006-07 round of the survey as it contains the additional variables required for this analysis.

In order to analyse the effect of higher rainfall on firms owned by members across different castes, we estimate the following equation:

$$log(y_f) = \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_d \times caste_i + \delta_d + \delta_F + \epsilon_i \quad \text{(B.32)}$$

First, we estimate the effect of higher rainfall on the following firm-level variables (y): (1) revenue, (2) material input, and (3) wages paid. This exercise provides a validation of whether our results from the panel data are prevalent in the cross-section as well.

Second, we estimate the effect on the following firm-level variables (y): (1) output product price, and (2) input product price. This exercise sheds light on two issues: to check whether the change in firms' revenues are driven by (i) a change in price or an actual growth in firm production, and (ii) a change in the quality of products, either sold or procured by these firms.

Third, we estimate the effect on the following firm-level variables (y): (1) all loans taken, (2) institutional loans taken, and (3) non-institutional loans taken. This exercise sheds light on two issues: to check whether the change in firms' outcomes in the previous regressions

Table B.5: Cross-sectional firm-level elasticities in rural India: MSME 2006-07

	(1) Revenue	(2) Material Input	(3) Wages	(4) Output Price	(5) Input Price	Loans (6) All	Loans (7) Institutional	Loans (8) Non-institutional	(9) log($mrpk$)
MC	-1.283***	-1.595***	-0.877***	-0.594*	-0.855*	-0.606	-0.612	0.266	-0.193
	(0.311)	(0.386)	(0.229)	(0.346)	(0.467)	(0.488)	(0.445)	(1.200)	(0.164)
LC	-1.497***	-1.973***	-0.897***	-0.321	0.089	-1.404*	-0.965	-0.141	0.219
	(0.315)	(0.385)	(0.232)	(0.352)	(0.435)	(0.716)	(0.630)	(2.493)	(0.238)
log(rainfall) × MC	0.133***	0.167***	0.096***	0.071	0.103	0.023	0.031	-0.093	0.033
	(0.045)	(0.056)	(0.034)	(0.050)	(0.066)	(0.068)	(0.064)	(0.174)	(0.023)
log(rainfall) × LC	0.131***	0.179***	0.071**	0.027	-0.063	0.109	0.055	-0.132	-0.019
	(0.046)	(0.057)	(0.034)	(0.052)	(0.063)	(0.105)	(0.092)	(0.361)	(0.034)
Observations	511,137	508,871	483,151	267,912	168,946	33,541	28,666	1,404	510,786
R-squared	0.423	0.458	0.469	0.358	0.393	0.425	0.482	0.678	0.33
District FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Columns (1) to (3) show the regressions of firm-level variables (logarithmic) on rainfall (logarithmic) in the year 2006-07. Columns (4) and (5) show the regressions of product-level prices on rainfall (logarithmic) in the year 2006-07. Columns (6) to (9) show the regressions of firm-level loans taken and $mrpk$ (logarithmic) on rainfall (logarithmic) in the year 2006-07. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

are driven by (i) a change in their borrowing from formal institutions, and (ii) a change in their borrowing from informal institutions such as the caste-network.

We also present a series of robustness checks in Section B.4.

Labour. In Table B.5, Columns (1) and (2), we begin by establishing that the results from the MSME panel shown in Table 3 hold in the cross-section as well. Secondly, Table B.5 shows that wages paid by LC firms also increase by 7.1% relative to HC firms, following higher rainfall, consistent with their constraints being relaxed and the material input purchase increasing. However, the table shows that the constraint of LC firms was especially relaxed with regards to their material input purchase than wages. This suggests that with increased demand, the constrained LC firms were able to obtain more *variable input* and necessary workers in order to expand.

Prices. To further explore the mechanism behind the effect of higher rainfall on firm revenue, we estimate the effect of higher rainfall on input and output prices. Table B.5, Columns (4) and (5) show that there is no significant difference in the average prices of LC firms relative to HC. Figure B.2 shows this pattern is present across the entire distribution of prices. First, considering input prices as a proxy for quality, we find no significant change in input prices, which shows a lack of evidence for the story of quality differences as an explanation for higher revenue. Also, with higher rainfall, there are no significant changes in output product prices. This result suggests that the quality of products produced also remained at similar levels. Secondly, these observations show that the relative increase in LC firms' revenue is driven by an increase in quantity produced and not prices. This evidence reassures us that the effect of rainfall on firm revenue, through its effect on factor prices is not a significant channel.

Credit. Thirdly, we investigate whether the positive income shift for LC households was transferred to LC firms through channels other than a shift in demand, namely that of loans. First, table B.5 shows that value of loans LC firms have taken is 68.5% lower than that of HC firms, in a district with median rainfall.[35] Second, with higher rainfall, there is no change in loans taken overall, from formal or informal institutions. This rules out any credit-side channel through which the LC income shift reaches firms, that is through formal (e.g., banks) or informal (e.g., caste networks) credit sources. An alternative outcome that may capture the generation of credit is the marginal revenue product of capital ($mrpk$). Any fall in this measure would suggest that firms have obtained funding (external or internal) and have invested in capital after rainfall. Table B.5, Column (9) shows no significant change along these lines. This evidence provides further support in the identification of cash-flow based constraints among firms in rural India.

Firm characteristics. In Table B.6, we interact rainfall with (i) the size of firms, and (ii) the nature of industry in which the firms operate.

One may be concerned that higher rainfall generates an alternative source of income for small firms, which might be driving their growth, instead of the caste-based demand channel. In that case, one would expect smaller firms to drive our baseline results in Table 3. Table B.6, Columns (1) and (2) show that this is not the case, and in fact, the effect of rainfall on firm growth is driven by LC firms above the median of revenue distribution.

Similarly, one may be concerned that the change in firms' outcomes are concentrated in some of the largest revenue-generating product-categories for LC firms. Table B.6, Columns (3) and (4) show a lack of any significant difference in the revenue, material input, and $mrpm$ levels between the top 20 product-categories for LC firms (by revenue) and other product-categories. It also shows that the effect of rainfall is not driven by these industries.

B.4 Robustness Checks

B.4.1 Robustness of wage patterns to worker sample

Table B.7 shows the estimated effects of rainfall on agricultural wages. They include observations of wages earned by workers of all ages. We find that the results are stable as in baseline Table 3, Columns (1) to (3), where the sample was restricted to workers between the ages of 18 and 60.

[35] Goraya (2023) shows that these differences are not driven by productivity, and establishes the existence of misallocation across caste due to credit-constraints.

Table B.6: Heterogeneity in firm-level elasticities in rural India: MSME 2004-07

Mediating variable (*hetvar*):	Small firms		Top LC products	
	(1)	(2)	(3)	(4)
Outcome:	Revenue	Material Input	Revenue	Material Input
MC	-1.415***	-1.771***	-2.161***	-2.016***
	(0.335)	(0.473)	(0.511)	(0.660)
LC	-1.772***	-2.298***	-1.956***	-2.701***
	(0.316)	(0.382)	(0.748)	(1.011)
log(rainfall) $\times$ MC	0.150***	0.190***	0.247***	0.208**
	(0.049)	(0.069)	(0.075)	(0.098)
log(rainfall) $\times$ LC	0.185***	0.245***	0.187*	0.287*
	(0.046)	(0.055)	(0.111)	(0.153)
hetvar	-1.593***	-2.097***	-1.043	-0.351
	(0.362)	(0.426)	(1.040)	(1.140)
log(rainfall) $\times$ *hetvar*	-0.029	0.008	0.141	0.052
	(0.052)	(0.062)	(0.153)	(0.168)
MC $\times$ *hetvar*	0.738*	1.135**	3.335**	2.508
	(0.415)	(0.547)	(1.538)	(1.910)
LC $\times$ *hetvar*	1.201***	1.629***	-0.128	-0.599
	(0.363)	(0.417)	(1.865)	(2.353)
log(rainfall) $\times$ MC $\times$ *hetvar*	-0.055	-0.107	-0.482**	-0.374
	(0.060)	(0.080)	(0.228)	(0.285)
log(rainfall) $\times$ LC $\times$ *hetvar*	-0.113**	-0.177***	0.044	0.101
	(0.053)	(0.061)	(0.277)	(0.341)
Observations	1,468,689	1,457,160	8,691	8,611
R-squared	0.549	0.551	0.538	0.562
District $\times$ Year FE	✓	✓	✓	✓
Product FE	✓	✓	✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall (logarithmic). Small firms indicate firms with below-median revenue, within caste-categories. Top LC products are the top 20 product-categories by total LC firm revenue. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

B.4.2 Robustness of consumption patterns to households' wealth status

Second, we check for the robustness of the effect of rainfall, across caste, to other determinants of household consumption patterns. Using land ownership as a proxy for within caste variation in households' illiquid wealth status, we analyse their consumption patterns by implementing a triple interaction exercise. This exercise addresses concerns regarding the reliance of our results to the segmentation of demand along the lines of wealth, instead of caste. The results reported in Table B.8 are consistent with the baseline results from Table 3, Columns (4) to (7). It further shows that LC consumers with less land, who would be more constrained, spend higher than LC consumers with high land. This shows that caste is an important determinant of consumption patters in the local economy, and within-caste wealth inequality is an additional determinant.

Table B.7: Wage elasticity in rural India for full sample of workers: NSS 2004-10

	(1) All	(2) Agri.	(3) Non-Agri.
log(rainfall)	0.050***	0.042*	0.067***
	(0.018)	(0.022)	(0.022)
MC	-0.012	-0.194	0.188
	(0.125)	(0.121)	(0.132)
LC	-0.131	-0.224**	-0.026
	(0.110)	(0.108)	(0.133)
log(rainfall) × MC	-0.016	0.025	-0.044**
	(0.019)	(0.018)	(0.019)
log(rainfall) × LC	-0.003	0.029*	-0.017
	(0.017)	(0.017)	(0.020)
Observations	142,534	70,221	72,312
R-squared	0.398	0.349	0.397
District FE	✓	✓	✓
Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes. The regressions of individual level wage (logarithmic) on rainfall (logarithmic) for the full sample of workers. *Agri.* stands for agricultural workers in rural areas. $Non - Agri.$ stands for workers employed in sectors non-agricultural sectors such as manufacturing, constructions, services, etc. in rural areas. The additional control variables are age, gender, education, land possessed, and crop season. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

B.4.3 Robustness of consumption patterns to households' educational status

In Table B.9, we interact rainfall with the educational qualification of the head of household, and with the number of meals per day consumed by the head of the household. In rural India, head of the household is responsible for a large portion of the household's liquid income. This exercise can help us understand whether the caste-driven consumption patterns documented in Table 3, Columns (4) to (7) can instead be explained by consumption patterns common to poor households. But this is not the case, as we see that the results are qualitatively similar to the baseline Table 3, Columns (4) to (7).

Table B.8: Household consumption elasticity in rural India (by caste), controlling for wealth status: NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
log(rainfall)	-0.059*	-0.029	-0.168*	-0.116
	(0.036)	(0.067)	(0.101)	(0.158)
MC	-0.876***	-0.397	-1.964**	-0.989
	(0.302)	(0.278)	(0.788)	(0.781)
LC	-1.302***	-0.465	-2.582***	-2.290***
	(0.260)	(0.312)	(0.731)	(0.726)
log(rainfall) × MC	0.105**	0.041	0.249**	0.130
	(0.043)	(0.041)	(0.116)	(0.117)
log(rainfall) × LC	0.153***	0.030	0.289***	0.316***
	(0.039)	(0.046)	(0.108)	(0.109)
land owned	-0.067*	-0.022	-0.193*	-0.025
	(0.037)	(0.037)	(0.105)	(0.112)
log(rainfall) × land owned	0.015***	0.018***	0.049***	0.027
	(0.006)	(0.006)	(0.016)	(0.017)
MC × land owned	0.081	0.036	0.188	0.102
	(0.052)	(0.052)	(0.134)	(0.132)
LC × land owned	0.113**	0.084	0.405***	0.258*
	(0.052)	(0.063)	(0.131)	(0.150)
log(rainfall) × MC × land owned	-0.012	-0.004	-0.026	-0.015
	(0.008)	(0.008)	(0.020)	(0.020)
log(rainfall) × LC × land owned	-0.018**	-0.011	-0.059***	-0.042*
	(0.008)	(0.010)	(0.020)	(0.023)
Observations	153,585	150,194	132,362	123,247
R-squared	0.365	0.329	0.261	0.262
District FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes. The regressions of households' monthly per capita consumption (logarithmic) on rainfall (logarithmic). Mfg. stands for consumption of clothing and footwear. The land variable is a categorical variable in some surveys and exact value in others. We synchronize the land variable across surveys. Yet, the average household's land holding across surveys are not consistent. Therefore, these results are to be interpreted with care. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

Table B.9: Household consumption elasticity in rural India (by caste), controlling for educational status: NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
log(rainfall)	-0.087**	-0.074	-0.059	-0.123
	(0.038)	(0.056)	(0.100)	(0.130)
MC	-0.567***	-0.454***	-1.343***	-0.861**
	(0.136)	(0.135)	(0.370)	(0.410)
LC	-0.990***	-0.695***	-1.669***	-2.218***
	(0.124)	(0.158)	(0.327)	(0.385)
log(rainfall) $\times$ MC	0.063***	0.052***	0.173***	0.112*
	(0.020)	(0.020)	(0.054)	(0.061)
log(rainfall) $\times$ LC	0.108***	0.063***	0.165***	0.283***
	(0.018)	(0.023)	(0.047)	(0.056)
education (head of HH)	-0.016	-0.037**	-0.056	0.026
	(0.013)	(0.018)	(0.039)	(0.041)
log(rainfall) $\times$ education (head of HH)	0.007***	0.011***	0.023***	0.005
	(0.002)	(0.003)	(0.006)	(0.006)
# of meals (head of HH)	-0.181*	0.213	0.338	0.101
	(0.098)	(0.168)	(0.280)	(0.341)
log(rainfall) $\times$ # of meals (head of HH)	0.021	-0.016	-0.041	0.000
	(0.015)	(0.025)	(0.042)	(0.051)
Observations	151,272	150,694	130,811	123,708
R-squared	0.409	0.288	0.286	0.241
District FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes. The regressions of household' monthly per capita consumption (logarithmic) on rainfall (logarithmic). Mfg. stands for consumption of clothing and footwear. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

B.4.4 Robustness of firm-level patterns to aggregate wealth status of households

Similarly, one may be concerned that the change in firms' outcomes are driven by consumption patterns of poor households, and not a result of caste. To alleviate this concern, we repeat the exercise from Table B.10, by controlling for land-owned, education, meals consumed by the head of household at the district$\times$caste$\times$year level, interacted with rainfall. We find that firms' outcomes in Table B.10, Columns (1) to (3) are qualitatively similar to the baseline Table 3, Columns (8) to (10). Further, Table B.10, Columns (4) to (14) show that the results in Tables B.5 are also robust to this concern.

Table B.10: Firm-level elasticities in rural India, controlling for aggregate wealth status

	MSME panel 2004-07		MSME cross-section 2006-07								
								Loans			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Revenue	Material Input	Revenue	Material Input	Wages	Output price	Input price	All	Institutional	Non-institutional	$\log(mrpk)$
MC	-1.650***	-2.057***	-1.560***	-1.963***	-1.045***	-0.867**	-1.236***	-0.768	-0.832*	0.032	-0.250
	(0.303)	(0.380)	(0.311)	(0.372)	(0.239)	(0.347)	(0.424)	(0.518)	(0.435)	(1.314)	(0.205)
LC	-1.890***	-2.473***	-1.636***	-2.204***	-0.978***	-0.588	-0.409	-1.404**	-0.910	2.370	0.239
	(0.314)	(0.394)	(0.323)	(0.393)	(0.253)	(0.370)	(0.444)	(0.711)	(0.663)	(3.273)	(0.279)
log(rainfall) × MC	0.188***	0.235***	0.172***	0.219***	0.120***	-0.048**	0.155**	0.049	0.063	-0.066	0.041
	(0.044)	(0.056)	(0.045)	(0.054)	(0.035)	(0.050)	(0.061)	(0.072)	(0.063)	(0.190)	(0.028)
log(rainfall) × LC	0.191***	0.255***	0.149***	0.210***	0.081**	0.063	0.000	0.112	0.046	-0.556	-0.028
	(0.046)	(0.058)	(0.047)	(0.057)	(0.037)	(0.055)	(0.065)	(0.104)	(0.097)	(0.485)	(0.039)
land owned	0.436	1.022	4.257***	5.726***	1.986**	2.875*	7.056***	-0.571	-0.757	-12.145	0.688
	(0.385)	(0.629)	(1.214)	(1.535)	(0.838)	(1.675)	(2.288)	(3.858)	(3.045)	(12.324)	(1.086)
log(rainfall) × land owned	-0.064	-0.156	-0.632***	-0.857***	-0.304**	-0.415	-1.059***	0.084	0.107	1.975	-0.069
	(0.060)	(0.099)	(0.189)	(0.241)	(0.129)	(0.266)	(0.346)	(0.591)	(0.469)	(1.964)	(0.162)
education (head of HH)	-1.294***	-2.091***	-3.047***	-4.206***	-1.747***	-3.439**	-6.366***	-3.368	-4.279*	22.398**	0.046
	(0.392)	(0.641)	(0.790)	(1.086)	(0.584)	(1.356)	(1.830)	(2.462)	(2.478)	(10.815)	(0.641)
log(rainfall) × education (head of HH)	0.200***	0.327***	0.455***	0.630***	0.268***	0.494**	0.887***	0.579*	0.673*	-3.546*	-0.205
	(0.059)	(0.096)	(0.117)	(0.161)	(0.085)	(0.192)	(0.256)	(0.348)	(0.348)	(1.792)	(0.091)
# of meals (head of HH)	0.097	0.053	-2.174**	-2.494*	-0.967	0.730	0.599	0.088	1.340	-57.933**	-1.125
	(0.632)	(0.757)	(1.015)	(1.286)	(0.688)	(1.959)	(1.607)	(3.617)	(2.911)	(24.596)	(0.969)
log(rainfall) × # of meals (head of HH)	-0.029	-0.028	0.335**	0.391*	0.150	-0.086	0.056	-0.144	-0.242	9.261**	0.139
	(0.092)	(0.113)	(0.161)	(0.205)	(0.107)	(0.295)	(0.251)	(0.534)	(0.433)	(4.057)	(0.145)
Observations	1,468,686	1,457,157	510,166	507,901	482,203	267,258	168,570	33,517	28,652	1,403	267,087
R-squared	0.422	0.461	0.424	0.458	0.469	0.358	0.394	0.425	0.483	0.682	0.311
District × Year FE	✓	✓	-	-	-	-	-	-	-	-	-
District FE	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Columns (1) and (2) show the regressions of firm-level variables (logarithmic) on rainfall (logarithmic) from MSME panel data 2004-07. Columns (4) to (6) show the regressions of firm-level variables (logarithmic) on rainfall (logarithmic) in the year 2006-07. Columns (7) and (8) show the regressions of product-level prices on rainfall (logarithmic) in the year 2006-07. Columns (9) to (10) show the regressions of firm-level loans taken and *mrpk* (logarithmic) on rainfall (logarithmic) in the year 2006-07. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

B.4.5 Robustness of firm-level patterns to product price level

In Table B.11, we restrict our attention to industries where the mean difference between the prices of LC and HC firms' products are less than 0.1 standard deviations. This exercises addresses concerns on whether the demand for low-quality/priced products from LC households is the mechanism behind our resutls. We find that firms' outcomes are qualitatively similar to the baseline Table B.5. This evidence adds support to our claim that caste-based consumer-firm linkages is key to understanding firm growth in rural India.

Table B.11: Firm-level elasticities in rural India, for industries with low price difference between LC and HC firms: MSME 2006-07

	(1) Revenue	(2) Material Input	(3) Wages
MC	-1.084***	-1.207***	-0.773***
	(0.250)	(0.359)	(0.183)
LC	-1.394***	-1.898***	-1.083***
	(0.264)	(0.373)	(0.201)
log(rainfall) $\times$ MC	0.124***	0.134**	0.093***
	(0.037)	(0.053)	(0.027)
log(rainfall) $\times$ LC	0.146***	0.203***	0.118***
	(0.039)	(0.056)	(0.030)
Observations	254,871	253,315	235,523
R-squared	0.442	0.509	0.570
District FE	✓	✓	✓
Product FE	✓	✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall (logarithmic). The sample is restricted to only those industries where the average product price difference between LC and HC firms are lower than 0.01 standard deviations. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

B.4.6 Robustness of firm-level patterns to alternative proxy for ethnic identity of firm

In our baseline empirical analysis, presented in Table 3, we measure the ethnic identity of the firm using the caste of the owner of the firm. It is also possible that the caste composition of workers (e.g., salespersons) employed in firms determine the perception of ethnic identity among consumers. In this exercise, we use the share of non-HC employees working in a firm as a measure of firms' ethnic identity. Table B.12 shows that firms which had a larger population of non-HC employees exhibited larger growth compared to other firms, after higher rainfall. This evidence is in line with our findings in our baseline analysis where after a higher rainfall consumers of LC and MC communities received a positive income shock and increased their expenditure on firms ethnically closer to their identity.

Table B.12: Firm-level elasticities in rural India, alternative proxy for firms' ethnic identity: MSME 2006-07

	(1) Revenue	(2) Material Input	(3) Wages
non-HC labour share	-0.552**	-0.565*	-0.499**
	(0.250)	(0.321)	(0.229)
log(rainfall) $\times$ non-HC labour share	0.070*	0.068	0.065*
	(0.037)	(0.047)	(0.034)
Observations	511,130	508,864	483,144
R-squared	0.407	0.443	0.458
District FE	✓	✓	✓
Product FE	✓	✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall (logarithmic). The sample is restricted to only those industries where the average difference between LC and HC firms are lower than 0.01 standard deviations. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

B.4.7 Robustness to alternative clustering of standard errors

In Table B.13, we cluster standard errors at the state-year level to allow for spatial correlation. The results are qualitatively consistent with our baseline results from Table 3.

B.4.8 Evidence on the transitory nature of rainfall deviations

Table B.14 shows that there is no serial correlation in rainfall across districts. This adds support to the interpretation of the exogenous variation in rainfall as having an unanticipated and transitory effect on the local economy.

Table B.13: Elasticities in rural India, alternative clustering of standard errors

	Individual wages			Household consumption				Firm outcomes	
	(1) All	(2) Agri.	(3) Mfg.	(4) All	(5) Mfg.	(6) Services	(7) Durables	(8) Revenue	(9) Material Input
log(rainfall)	0.050***	0.041*	0.069***	-0.005	-0.044	-0.002	-0.095	-	-
	(0.019)	(0.023)	(0.019)	(0.027)	(0.027)	(0.051)	(0.069)	-	-
MC	-0.150	-0.250**	-0.043	-0.583**	-0.440***	-1.357***	-0.915	-1.443***	-1.750***
	(0.120)	(0.136)	(0.130)	(0.228)	(0.158)	(0.339)	(0.556)	(0.423)	(0.509)
LC	-0.150	-0.250**	-0.043	-1.015***	-0.677**	-1.706***	-2.371***	-1.763***	-2.273***
	(0.115)	(0.106)	(0.133)	(0.266)	(0.268)	(0.533)	(0.620)	(0.373)	(0.474)
log(rainfall) × MC	-0.014	0.029	-0.040**	0.058*	0.042*	0.155***	0.107	0.158**	0.191**
	(0.018)	(0.020)	(0.019)	(0.033)	(0.023)	(0.050)	(0.083)	(0.061)	(0.074)
log(rainfall) × LC	-0.001	0.033**	-0.015	0.100**	0.048	0.138*	0.286***	0.172***	0.225***
	(0.018)	(0.016)	(0.020)	(0.040)	(0.039)	(0.080)	(0.091)	(0.054)	(0.069)
Observations	131,991	63,531	68,459	154,241	153,586	132,873	126,125	1,468,689	1,457,160
R-squared	0.401	0.353	0.394	0.351	0.251	0.225	0.21	0.422	0.460
Controls$_{it}$	✓	✓	✓	-	-	-	-	-	-
District FE	✓	✓	✓	✓	✓	✓	✓	-	-
Year FE	✓	✓	✓	✓	✓	✓	✓	-	-
District × Year FE	-	-	-	-	-	-	-	✓	✓
Product FE	-	-	-	-	-	-	-	✓	✓

Notes. In Columns (1) to (3), the regressions are of individual-level daily wage (logarithmic) on rainfall (logarithmic). $Agri.$ stands for agricultural workers in rural areas. $Non - Agri.$ stands for workers employed in sectors non-agricultural sectors such as manufacturing, constructions, services, etc. in rural areas. The additional individual-level control variables included are age, gender, education, land possessed, and crop season. Sample includes individuals between the ages of 18 and 60. In Columns (4) to (7), the regressions are of households' monthly per capita consumption (logarithmic) on rainfall (logarithmic). $Mfg.$ stands for consumption of clothing and footwear. In Columns (8) to (10), the regressions are of firm-level variables (logarithmic) on rainfall (logarithmic). Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at state-year level in all regressions, *** p<0.01, ** p<0.05, * p<0.1.

Table B.14: Testing for serial correlation in rainfall

	MonsDev$_{d,t}$ (1)	MonsDev$_{d,t}$ (2)
MonsDev$_{d,t-1}$	0.008	-0.015
	(0.034)	(0.041)
MonsDev$_{d,t-2}$		-0.019
		(0.025)
Observations	8,330	7,840
R-squared	0.197	0.213
District FE	✓	✓

Notes. This table tests for serial correlation in rainfall.

$$MonsDev_{d,t} = \alpha + \beta_1 MonsDev_{d,t-1} + \beta_2 MonsDev_{d,t-2} + \delta_d + \epsilon_{dt}$$

MonsDev is the deviation in monsoon rainfall from the median monsoon rainfall in a district since 1979. The unit of observation is district-year and the outcome sample includes rainfall between the years of 1999 and 2016. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.