Identity, Market Access, and Demand-led Diversification*

Sampreet Singh Goraya† Akhil Ilango‡

Abstract

Using Indian microdata on employer-employee caste composition and household consumption, we document demand segmentation along caste lines, limiting firms' market penetration and reducing firm size in the economy. We develop a model where consumers prefer goods produced by socially closer groups, and firms overcome these barriers by hiring employees from the target consumer group. We identify the structural parameters governing demand segmentation using rainfall-induced demand shocks. Counterfactuals indicate that social identity-driven barriers restrict the growth of high-quality firms while sustaining low-quality ones. Lowering the cost of hiring out-group employees expands firm size by improving market access and enhances consumer welfare through greater variety of products.

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

JEL classification: *O11, L11, L25, M14*

*This Draft: July 20, 2025. We thank our discussants, Aruni Mitra, Manaswini Rao, and Sophie Nottmeyer for insightful comments. We thank Murat Alp Celik, Roland Benabou, Fabio Blasutto, Shoumitro Chatterjee, Ritam Chaurey, Antonio Ciccone, Klaus Desmet, Pascaline Dupas, Tore Ellingsen, Farid Farrokhi, Jonas Hjort, Suzanna Khalifa, Ryan Kim, Gueorgui Kambourov, Tatjana Kleinberg, Pravin Krishna, Sylvie Lambert, Aruni Mitra, Kaivan Munshi, Eduardo Morales, Martina Bjorkman Nyqvist, Stephen Redding, Martin Rotemberg, Karthik Sastry, Manuel Garcia-Santana, Carolina Villegas-Sanchez, Mine Z. Senses, Garima Sharma, Abhijeet Singh, Joseph Steinberg, Maria Micaela Sviatschi, Michèle Tertilt, and Eric Verhoogen as well as the seminar participants at the NYU, PSE, Mannheim, University of Toronto, Princeton, John Hopkins University, UCSD, ULB, Turku, UC Merced, Ashoka University and conference participants at CEPR Manchester, STEG CEPR, SED, CEPR Firms, Labour Markets, and Development Workshop, Networks and Development Workshop Naples, SITES, ISI, UNU-WIDER, SEED, and APIOC for helpful discussions and comments.

†Stockholm School of Economics. Email: sampreetgrg@gmail.com

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

1 Introduction

Recent evidence makes it clear that buyers care about who they buy from. In high-income settings, both experimental and observational studies show that seller identity shapes demand—see Doleac and Stein (2013), Davis, Dingel, Monras, and Morales (2019), Alsan, Garrick, and Graziani (2019), Gil and Marion (2022) and Takeshita, Wang, Loren, Mitra, Shults, Shin, and Sawinski (2020).

Developing-country evidence points in the same direction. Scheduled-caste households in India purchase more grain when the vendor is also scheduled-caste Nagavarapu and Sekhri (2016); customers in Sub-Saharan Africa shy away from sales agents of certain genders Kelley et al. (2024); and stock-market participants trade differently when a firm's CEO shares their ethnic background Hjort et al. (2021).

Taken together, these findings imply that identity-based segmentation can splinter markets and hinder scale economies, yet we know little about the macroeconomic consequences. This paper fills that gap. We examine how firms navigate fragmented demand, document the hiring strategies they deploy to broaden market access, and quantify the aggregate costs of identity-driven barriers.

We emphasise that the *social distance* between consumers and firms is a significant barrier to market integration in developing countries.[1] To this end, we develop a novel framework where consumers prefer goods produced by socially closer groups and calibrate its main parameters using microdata on employer-employee ethnic composition and household consumption.

In the model, firms can overcome these barriers – reduce social distance – by hiring employees from the target consumer group. However, firms face a trade-off: becoming closer to the socially distant group may make them less attractive to their current consumers. We provide two main results: (i) Large firms are more likely to sell to a diverse set of customers and thus have a diverse workforce composition, a phenomenon we label as demand-led diversification. (ii) The workforce composition within a firm

[1] Firm identity plays a salient role, especially in low-income countries. Recently, the government of Uttar Pradesh – the largest state in India – issued a directive requiring certain shops and eateries to display the names and identities of their owners and employees (Times of India, July 2024). Nagavarapu and Sekhri (2016) show that scheduled Castes (LC) are more likely to buy grains when facing LC sellers. Kelley et al. (2024) documents customer discrimination associated with sales agents' gender in Sub-Saharan Africa. Hjort et al. (2021) shows that CEO identity is important in stock trading.

is shown to be proportional to the local distribution of demand between social groups of consumers.

We test the model's predictions in the context of rural India, where the majority of the Indian population resides. It is a setting where caste norms, which fragment the population into groups, are strong, and firms are small. Caste is inherited at birth and determines one's social identity. Historically, the caste system restricted inter-caste interactions and promoted discriminatory practices towards low-ranked castes (LC, hereafter). Despite rapid socioeconomic changes in recent decades, caste remains an important and salient feature of Indian society (Munshi, 2019).

We use a unique dataset — 2006-07 Micro, Small, and Medium Enterprises (MSME) Survey — that incorporates caste information for employers and employees, as well as product revenues and quantities (Goraya, 2023).[2] The data allows us to create a balanced panel of three years for revenues and material costs.

Empirical Facts on the Firm-side. Our analysis uncovers *three* key patterns. First, hiring is strongly homophilic: roughly 75% of employees share their employer's caste, with the own-caste share falling from about 85% in the smallest decile of firms to around 55% in the largest. Second, this drop is far steeper in customer-facing industries– where worker-client interaction is most visible.[3] Third, the reduction in homophily is notably larger for firms owned by low-ranked castes than for those headed by high-ranked castes.

Empirical Facts on the Consumer-side. Next, we document identity Engel curves, the sensitivity of consumer demand to perceived social distance of a product. We define social distance as the relative share of employees of a certain caste in a product market. We find that as the (high-ranked castes) HC's relative employee share in a product market increases, HC consumer expenditure increases relative to the LC consumer while controlling for income and wealth. We repeat this exercise for different pairs of castes and find similar results, suggesting a strong association between consumer demand and perceived product identity.

Empirical Identification Strategy. However, these correlations are potentially driven

[2]MSME survey is the only nationally representative dataset of Micro, Small, and Medium Enterprises that provides information on employee's social group. This combined with the information of financial variables and product prices makes it a unique fit for our study.

[3]Customer-facing sectors include firms in Repair & Maintenance, Wholesale & Retail, and selected other services. They account for roughly 30–40 percent of all firms in our data.

by omitted variables. To establish causation, we use local rainfall shocks that have been widely used as demand shocks. Following Jayachandran (2006), we divide districts with negative (-1), neutral (0), and positive (+1) rainfall shocks. Crucially, in the case of India, we find that these shocks asymmetrically affect certain castes, especially LC households.

Next, we examine how rainfall shocks affect households' Monthly Per-Capita Expenditure (MPCE), using district $\times$ year and caste fixed effects. A positive rainfall shock leads to an average 8.6% increase in MPCE for LC households compared to HC households, particularly in non-food categories like footwear, clothing, services, and durables. We find that these results are robust to controlling for wealth or education differences across consumer groups; further establishing the role of caste as the driving force behind this effect.

To test how LC demand shocks transmit to firms, we estimate the contemporaneous effect of rainfall shock on firms, employing district $\times$ year, sector or product $\times$ year, and caste fixed effects. We find that a positive rainfall shock increases the revenue of LC-owned firms by 13.4%, on average, relative to HC-owned firms. We find that the effect is concentrated among products that (i) witness the highest LC household demand increase, (ii) have a low dependence on agricultural inputs, and (iii) relatively large LC-owned firms. We find no evidence of these firm-level effects being driven by sector specialization (that is, in some niche sectors where LC-owned firms dominate), or by low-quality goods.[4] Next, we evaluate the role of information friction in consumer markets – where it is difficult to differentiate between high and low-quality products. We employ two proxies: (i) Rauch (1999) classification of product differentiation and (ii) product price dispersion. Our results are similar even in relatively homogeneous product markets, where quality is easier to assess. This suggests that preference for own-caste products is more likely to be driving the results.

In assessing workforce adjustments to demand shocks, we are limited to one cross-section, thus we exploit district, sector, and caste fixed effects together with rainfall

[4]We measure product quality by prices of the raw material used in production. We find that the effects are equally strong in product markets where LC- and HC-owned firms produce goods with similar quality. Further, using the information on village-level population shares, we show that potential geographic segregation by castes is not driving the results. Lastly, we examine if rainfall shocks loosen financial frictions for LC-owned firm owners, we find no systematic relationship between the amount of informal and formal loans, and the marginal revenue product of capital—a widely used proxy for capital market frictions in the misallocation literature—and rainfall shocks in our data.

shocks for our identification. We find that a positive rainfall shock, contemporaneous to the above effects, decreases HC employees' share and increases LC employees' share among LC-owned firms. Further, we exploited heterogeneity in product market competition to test how HC-owned firms respond to changes in local demand.[5] We find that HC-owned firms increase their share of LC employees in more competitive markets in response to an increase in the demand share of LC consumers. Finally, we highlight that our results are unlikely to be driven by an increase in LC households' labour supply. Although the rainfall shock affects the whole district, the increase in LC employees among HC-owned firms is concentrated in customer-facing sectors, where employee's caste holds more significance. Cumulatively, this evidence suggests that local demand composition influences a firm's employee composition.[6]

Micro-to-Macro. We calibrate the model using causally identified elasticities to understand the effect of caste-led demand fragmentation on the macroeconomy. This exercise hinges on calibrating four parameters: the taste for identity in consumption, fixed operational costs, fixed trading costs across castes, and the scale parameter of quality 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 asymmetrically impacts the LC caste group. Specifically, within our framework, we calibrate the taste for identity parameter by matching the revenue elasticity of LC-owned firms, relative to HC-owned firms.[7] For the remaining parameters, we match the firm sales distribution and the share of firms hiring cross-caste.

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 the social distance between castes. Using this calibrated model, we show that taste for identity makes firms' markets socially local. Firms are less likely to trade with socially distant castes, less likely to hire

[5]HC-dominated markets are defined by product-district pairs where the share of firms/revenues/employees for HC-owned firms is above a certain cut-off.

[6]One concern may remain that firms, irrespective of the nature of the temporary shock, always expand by hiring more LC employees. To make progress, we use foreign demand shocks – that are caste neutral – as an alternative source of firm growth, we find no changes in the workforce caste composition among firms. The LC employee share remains constant among HC-owned firms. Thus, firms are unlikely to consistently hire LC employees when experiencing any temporary shocks. While rainfall shocks, by changing the local demand composition across castes, induce changes in workforce composition.

[7]Due to data constraints, we restrict ourselves and estimate a single taste elasticity, thus abstracting away from differences across castes.

employees from other castes, and tend to remain small. We find empirical support for this: regions that are populated by fewer castes are associated with larger firms and this relationship is stronger in customer-facing sectors.

In a series of counterfactual experiments, we find that preferences for in-group firms significantly impact the aggregate economy. Doubling these barriers reduces aggregate income by approximately 5%.[8] While these preferences constrain aggregate income in our setting, they may also yield certain benefits in other circumstances. Therefore, a practical implication of these results is that the cost of policies aimed at addressing these barriers should not exceed the highlighted gains.

Next, we vary the cost of cross-caste hiring while keeping the preference for in-group consumption constant. This exercise is motivated by the vast regional disparities in cross-caste hiring. Our results demonstrate that lowering these costs effectively reduces barriers to market access. As a result, firms are encouraged to cater to socially distant castes, leading to reduced homophily throughout the economy. This shift also prompts firms to enter diverse markets, expanding consumer choice, and enhancing overall welfare. This raises the question: Can policies influence the cost of cross-group hiring and promote trade? Recent studies suggest that inter-group interactions can foster positive social attitudes and mitigate biases. Therefore, short-term subsidies aimed at promoting workforce diversity may offer a viable strategy.[9]

Literature Review. First, this paper relates to the literature on the importance of demand-side drivers for firm size (Foster et al., 2008, 2016; Startz, 2016; Hardy and Kagy, 2020; Einav et al., 2021; Bernard et al., 2022; Bold et al., 2022; Vitali, 2022; Bassi et al., 2023 and Tan and Zeida, 2023).[10] Hottman et al. (2016) and Eslava et al. (2024) show that product appeal explains most of the variation in firm size. This paper takes a step forward and highlights one channel that constitutes product appeal – the identity of the firm.[11] The demand segmentation by groups limits the scale economies.

[8] This result is in line with the literature that documents negative effects of social fragmentation on economic development; Alesina et al., 1999; Alesina and La Ferrara, 2000, 2005, and Montalvo and Reynal-Querol, 2005, 2021.

[9] European countries have implemented diversity promotion policies (see OECD, 2020). The potential of a diverse workforce to expand customer bases has been discussed in Hunt et al. (2018).

[10] Also, see Allen (2014); Donaldson (2015); Atkin and Donaldson (2015); Jensen and Miller (2018) and Asturias et al. (2019) for evidence on large geographical trading barriers.

[11] Recent evidence suggests that customers care about seller identity in advanced economies. See, for instance, Doleac and Stein (2013); Davis et al. (2019); Alsan et al. (2019) and Takeshita et al. (2020).

This helps us better understand why firms tend to be small in developing countries (Hsieh and Olken, 2014; Hsieh and Klenow, 2014 and Bento and Restuccia, 2017).

Second, this paper is related to the literature on ethnic networks, preferences, and trade (McCallum, 1995; Rauch, 2001; Rauch and Trindade, 2002; Combes et al., 2005; Felbermayr and Toubal, 2010; Michaels and Zhi, 2010; Aker et al., 2014; and Desmet et al., 2023). Specifically in India, Anderson (2011), and Emerick (2018) show that caste hinders the trade of groundwater and seeds, respectively. Fujiy et al. (2022) and Boken et al. (2022) find a positive association between caste and firm-to-firm trade. However, there is limited evidence on consumption homophily by caste. We fill this gap by documenting identity Engel curves, and by providing causal evidence of an asymmetric transmission of caste-specific demand shocks to firms.[12] Further, we uncover a novel channel through which demand-side tastes may spillover to labour markets.

Third, this paper contributes to the literature that links the ethnic/religious composition of the labour force with productivity. Hjort (2014); Afridi et al. (2020), and Ghosh (2024) show that ethnic heterogeneity lowers firms' productivity, whereas Bhagavatula et al. (2018) and Brinatti and Morales (2021) find that better performance and worker diversity are positively linked. However, due to data limitation, so far there is no evidence on how an employer's caste matters for employee composition and scale in the Indian organized sector. We document novel facts regarding entrenched homophily in hiring and how it depends on the nature of the sector. Finally, we highlight a new channel – demand led diversification – highlighting that more ethnically diverse firms are larger because they can appeal to diverse consumer markets.

Fourth, this paper is connected to the literature that studies trade across interdependent markets (Albornoz et al., 2012; Schmeiser, 2012; Chaney, 2014; Morales et al., 2019; and Alfaro-Urena et al., 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, that is, demand spillovers.

This paper is organized as follows. Section 2 describes the Caste system. Section 3 presents a quantitative framework. Section 4 introduces the data. Section 5 discusses the empirical strategy and documents the results. Section 6 outlines the quantitative analyses. Section 7 concludes. All proofs are deferred to the Appendix A.

[12] See also Adhvaryu et al. (2013); Chaurey (2015); Kaur (2019); Santangelo (2019); and Gupta (2020) for recent work on the effects of rainfall on the Indian labour market.

2 Institutional Background

The caste system in India is a form of social stratification. Caste is assigned at birth, propagated through endogamy, and has dictated social interactions and exclusionary practices.[13] Historically, it divided individuals into different occupations and enforced the social ordering of these groups based on purity and prestige, thereby underpinning social identity. In descending order of caste purity, the hierarchy is as follows: Brahmins (priests and teachers), Kshatriyas (rulers and soldiers), Vaishyas (merchants and traders), and Shudras (labourers and artisans). Additionally, two groups fall outside and below the caste system: the first group, known as Dalits (referred to as Scheduled Castes (SC) since 1947), and the second group, known as Scheduled Tribes (ST). These groups were positioned at the bottom of the hierarchy, left to do menial tasks and subjected to discrimination and exclusion.

More generally, the caste hierarchy continues to shape social interactions in modern India, including the language used in conversations, the sharing of food, and marriage practices. In the post-independence era, the Indian government recognized the longstanding caste discrimination and instituted affirmative action programs for three "backward" groups, categorized as scheduled castes (SCs), scheduled tribes (STs), or other backward classes (OBCs).

Academics have extensively documented the importance of traditional occupations.

Critics of the caste system emphasize that the hierarchical and occupational nature of caste are deeply intertwined. Most high-wage occupations were historically forbidden to lower castes, and certain occupations, such as cooking, have become linked to upper castes due to their inherent association with ideals of purity. Conversely, occupations associated with lower castes, such as waste removal, leather work, or agricultural labor, tend to offer low returns, are unpleasant or servile, and have become stigmatized for higher castes

Within these broader groups, there are thousands of sub-groups known as "Jatis." Several papers have highlighted the effects of cultural proximity on a range of economic outcomes both at the broader caste level and at the Jati level. For instance, cul-

[13] Bidner and Eswaran (2015) describe 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.

tural proximity among caste members may reduce transaction costs in contexts with severe informational and contractual frictions (see Fisman et al., 2017 for loan outcomes and Munshi and Rosenzweig, 2016 for insurance). However, the existence of caste networks could also lead to preference-based transactions within the caste, resulting in discrimination and resource misallocation (see Banerjee and Munshi, 2004; Goraya, 2023 for capital misallocation, Anderson (2011); Boken et al. (2022) for trading frictions).[14] Similarly, the bias towards in-group products may represent both positive aspects – caste alleviating transaction costs associated with information frictions related to product features and quality, or caste may have negative aspects – discrimination against out-group products. We parse out these two mechanisms by exploiting the degree of product differentiation – a widely used measure in the trade literature in Section 5.3.

Why would consumers care about the caste of the firm owner or the employee composition of firms? As mentioned above, the caste system restricts cross-group interactions. For instance, Lowe (2021) finds significant caste-based homophily, with participants being twice as likely to form connections with someone from the same caste compared to others. Consequently, firms that predominantly employ low-ranked castes may face lower demand from high-ranked castes' consumers, who may anticipate interacting with low-ranked castes employees during the transaction. Therefore, caste is likely to be more important in sectors where employee-customer interactions are crucial. We will analyse these sectoral differences in Section 5.2.3.

For the remainder of the paper, we focus on a broader classification of the caste system, following Indian administrative practices, due to data constraints. Historically disadvantaged castes are denoted as "LC," which includes Scheduled Castes and Scheduled Tribes. Middle-ranked castes are denoted as "MC," which includes Shudras (also known as Other Backward Castes, OBC, falling between the traditionally dominant upper and disadvantaged lower caste categories). Individuals belonging to historically privileged castes are denoted as "HC." Our objective is to evaluate how cultural proximity across these broad groups affects trade and hiring practices. As previous work has shown that the caste system operates even at the Jati level and pro-

[14] See also other contributions on labour markets (Madheswaran and Attewell, 2007; Oh, 2023; Cassan et al., 2021), health and education (Munshi and Rosenzweig, 2006; Hnatkovska et al., 2012, 2013; Spears and Thorat, 2019), preferences (Atkin et al., 2021), and gender norms (Agte and Bernhardt, 2023).

motes firm-to-firm trade within Jatis (see Boken et al., 2022 and Fujiy et al., 2022), our results, discussed in the next section, may underestimate the effect of caste on trade between consumers and firm owners, as we assume friction-less trade within broader caste-groups such as HCs, MCs and LCs.

3 Theoretical Framework

In this section, we describe a model of within-district trade. There are $\mathcal{S}$ groups in the district and the ethnic distance between two groups 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 social distance and $d_{ss'} = 1$ implies the socially most distant groups. We now describe the household sector and the production sector.

3.1 Households

For each group, there is a representative household s. It has a labour endowment of L_s. The household has log preferences over two types of goods: homogeneous goods and differentiated varieties. It solves the following problem

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

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

$$I_s = w_s L_s + \Pi_s \geq P_H C_{H,s} + \sum_{\Omega_s} p(z(\omega), s, s') c(z(\omega), s, s'), \tag{3}$$

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 groups at a social distance $d_{ss'}$. The households have CES preferences over endogenous Ω_s differentiated varieties, where σ denotes the elasticity of substitution between varieties. The $c(z(\omega), s, s')$ denotes the units of consumption by household s of good quality $z(\omega)$ and produced by group s', $q(z(\omega), s, s')$ denotes the utility derived per unit good consumed, and $p(z(\omega), s, s')$ denotes the price. The total income of the household is denoted by I_s which is composed of total wage income and net profits of the firms that belong to the same group. The optimal expenditure and the demand for a variety is given by an iso-elastic demand curve:

$$c(z(\omega), s, s') = y(z(\omega), s, s') = q(z(\omega), s, s')^{\sigma} p(z(\omega), s, s')^{-\sigma} \kappa_s, \tag{4}$$

where $\kappa_s = C_{D,s} P_{D,s}^{\sigma} = (1-a) I_s P_{D,s}^{\sigma-1}$ denotes the market demand. A variety's appeal $q(z(\omega), s, s')$ is given by

$$q(z(\omega), s, s') = z(\omega) \Psi(d^*_{z,s,s'}),$$

where, $\Psi(d^*_{z,s,s'})$ captures the effect of social distance on product appeal, and $\widetilde{d}^*_{z,s,s'}$ is the optimal social distance. The taste shifter satisfies the following properties: $\frac{\partial q}{\partial z} > 0$, the taste is higher for high quality goods and $\frac{\partial q}{\partial d^*_{z,s,s'}} < 0$, the taste is lower for goods sold by out-group firm.

3.2 Production Sector

3.2.1 Homogeneous good sector

There is a representative firm of each group 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 labour producing 1 unit of homogeneous good. Its price is set equal to one, so that if group s produces this good, the wage is normalised to one. We assume that the utility from homogeneous good is sufficiently important for households, that is a is large enough, such that all groups produce the homogeneous good and all wages are equal to one.[15]

3.2.2 Differentiated Good Sector

We assume that differentiated products are produced by a continuum of firms. Firms use Cobb-Douglas production technology with constant returns to scale, $y(z, s) = \ell_s$. The labour is the only input in production and it is from the same group as the firm owner. There is is fixed cost of production f^d_{ss}. As the marginal cost of production is constant, firms maximise profits in each market. To access out-group markets, firms

[15] Real wage will be different across groups, with LC being the poorest. It is straightforward to allow for different nominal wages across groups.

need to pay a fixed cost of $f^d_{ss'}$ every period.[16] Given consumers' demand, a firm with product quality z, and from group s solves the following profit maximisation in each ethnic market at a cultural distance $d_{ss'}$.

$$\pi(z,s',s) = \max_{p(z,s',s),\ell(z,s)} p(y(z,s',s))y(z,s',s) - \mathcal{C}(s)y(z,s',s) - f^d_{s',s} \tag{5}$$

$$s.t. \quad y(z,s',s) = q(z,s',s)^{\sigma} p(z,s',s)^{-\sigma}\kappa_{s'} \tag{6}$$

The first-order condition gives us the standard results $p(z,s',s) = \frac{\sigma}{\sigma-1}\mathcal{C}_s$. There is no price discrimination. As usual, markup is decreasing with the elasticity of substitution. We divide the firm's problem into two stages. First, firms solve for optimal social location. Second, given their social location, they solve for optimal size. We solve the model backward.

3.2.3 Firm Scale

At the extensive margin, firms only sell to an out-group if the profits are higher than the fixed cost of selling to that group. At the intensive margin, sales to each group depend on two opposite forces. First, socially distant groups have lower appeal; thus, the sales to culturally distant groups are lower. However, demand increases with the size of the out-group consumers. Let us define a set $\mathcal{M}$ that contains all possible combinations of markets and $\mathcal{J}_s \in \mathcal{M}$ be the set of the active markets, Y_s and L_s as the total output sold and labour employed for the firm s. The firm revenue is

$$R(z,s) = \sum_{s'\in\mathcal{J}_s} r(z,s',s) = \left(\frac{\sigma}{\sigma-1}\mathcal{C}_s\right)^{1-\sigma} z^{\sigma} \sum_{s'\in\mathcal{J}_s} \Psi^{\sigma}_{z,s',s}\kappa_{s'}. \tag{7}$$

3.3 Endogeneous Group Proximity

We begin by further elabourating on the objective function of firms. The firms can position themselves closer or farther away from certain groups. This feature allows them to overcome the disadvantage of being born into a small group. However, the incentive to be closer to certain groups to take advantage of their large size comes at the cost of losing demand from in-group consumers – *demand spillovers*. Additionally, there are

[16] The presence of fixed cost is motivated by the fact that the probability of trading within-group is double that of trading outside the group. See, for instance, Boken et al. (2022) and Fujiy et al. (2022).

costs associated with changes in firms' identity, as parameterized below. These costs are paid in wages – firms can hire employees from out-groups to interact with their consumers. All groups are placed in $\mathcal{R}_+^{\mathcal{N}}$ space, where $\mathcal{N} = \mathcal{S}$.

A column vector, $\mathcal{X}_s = \{\mathcal{X}_{s,1}, ..., \mathcal{X}_{s,\mathcal{N}}\}$, contains the Cartesian coordinates of group s. The Euclidean distance between two groups represented by $\mathcal{X}_s$ and $\mathcal{X}_{s'}$ is $d^2_{s,s'} = \sum_{k=1}^{\mathcal{N}} (d_{s,s',k})^2$, where $d_{ss',k} =\mid \mathcal{X}_{s,k} - \mathcal{X}_{s',k} \mid$ is the L1 distance in the k^{th} dimension. Further, let us define the distance (relative to her initial position) moved by a firm owner born in 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 $\Delta d^2_s = \sum_{k=1}^{\mathcal{N}} (\Delta\mathcal{X}_{s,k})^2$. We assume that firms pay a group-dependent moving cost that is given by a function $\Phi(\Delta\mathcal{X}_s; \Gamma_s)$, where Γ_s is a $\mathcal{N} \times 1$ column vector of parameters that disciplines the costs of moving. Here, we allow for the possibility that LC-owned firms may pay higher to move closer to HC consumers than HC-owned firms pay to move closer to LC consumers. Profits of the firm owner s is given by

$$\Pi_D(z, s, \Delta\mathcal{X}_s) = B_{z,s} \sum_{s' \in \mathcal{J}_s} \Psi(d_{s',s}, \Delta d_{z,s})^{\sigma} \kappa_{s'} - \Phi(\Delta\mathcal{X}_{z,s}; \Gamma_s) - \sum_{s' \in \mathcal{J}_s} f^d_{s',s}, \tag{8}$$

where $B_{z,s} = \frac{1}{\sigma} \left(\frac{\sigma}{\sigma - 1} \mathcal{C}_s \right)^{1-\sigma} z^{\sigma}$. We assume that the identity taste shifter Ψ takes the form

$$\Psi_{z,s',s} = e^{-\hat{\beta}\left(\sum_{k \in \mathcal{N}} (d_{s',s,k} - \Delta\mathcal{X}_{z,s,k})^2\right)}, \tag{9}$$

where taste depends on the social distance between the firm and consumer, that is $d^*_{s',s,k} = d_{s',s,k} - \Delta\mathcal{X}_{s,k}$. The parameter $\hat{\beta}$ denotes the taste elasticity. A first-order Taylor approximation of this function gives us $\Psi^{\sigma} \approx 1 - \beta \left(\sum_k (d_{s',s,k} - \Delta\mathcal{X}_{s,k})^2 \right)$ with $\beta = \hat{\beta}\sigma$. This gives us a quadratic taste function that is reminiscent of Hotelling (1929) quadratic cost functions. The advantage of working with the quadratic taste function is that it gives us a closed-form solution to the optimal social location problem.

3.3.1 Firm's Optimal Social Location and Employee Caste Composition

We now solve for the closed-form solution to the optimal social location problem. A firm's original social location or identity is the same as its owner's identity. In our context, choosing optimal social location is akin to deciding how much a firm wants to sell to a social group: the closer a firm is to a group, the higher the demand it faces from that group. The costs paid are in terms of the labour of the targeted group, as they change the labour composition of the firm.

For instance, consider a case with three groups in three dimensions (X,Y,Z) whose coordinates are given by: LC (1,0,0), MC (0,1,0), and HC (0,0,1). They may be represented in the $\mathcal{R}^3_+$ space, as in Figure 1. The cost of moving $\tilde{\gamma}_{z,s,k}$ is proportional to the size of the firm which depends on its quality z, on its initial location s, and the dimension along which it desires to move k. In particular, $\tilde{\gamma}_{z,s,k} = \gamma z^{\sigma}$, where γ is a cost shifter constant across groups.[17] The γ captures the firm's cost of diversifying the workforce beyond the prevailing wage of out-group workers, including factors like distaste and potential workforce conflicts. In the baseline model, we assume that $\gamma = 1$ is constant across groups; however, it can be extended to scenarios where lower group (LC) groups face higher costs when hiring higher caste (HC) individuals, and vice versa. The total moving cost for a firm is given by [18]

$$\Phi(\Delta\mathcal{X}_{z,s};\Gamma_{z,s}) = \sum_k \gamma_{z,s,k}\Delta\mathcal{X}^2_{z,s,k}, \quad \text{where} \quad \gamma_{z,s,k} = \gamma z^{\sigma} \tag{10}$$

Using Equations (8), (9) and (10), we can rewrite the firm problem as

$$\Pi_D(z,s,\Delta\mathcal{X}_{z,s}) = \max_{\mathcal{J}_s,\Delta\mathcal{X}_{z,s}\geq \underline{X}_{z,s}} B_{z,s}Y\left(1 - \sum_{\mathcal{J}_s}\sum_k \lambda_{s',s,k}(d_{s',s,k} - \Delta\mathcal{X}_{z,s,k})^2\right) - \sum_{s'\in\mathcal{J}_s} f^d_{s',s}, \tag{11}$$

where $Y = \sum_s \kappa_s$, and $\lambda_{s',s,k} = \frac{\beta\kappa_{s'} + B^{-1}_{z,s}\gamma_{z,s,k}\mathbb{1}_{s=s'}}{Y}$. The above problem gives us $\mathcal{J}^*_s$, the set of active markets, and the optimal distance $\Delta\mathcal{X}^*_{z,s}$. The distance moved towards an open out-group market (for which the fixed cost is paid) should be bigger that $\underline{X}_{z,s',s} = f^d_{s',s}/z^{\sigma}$.[19]

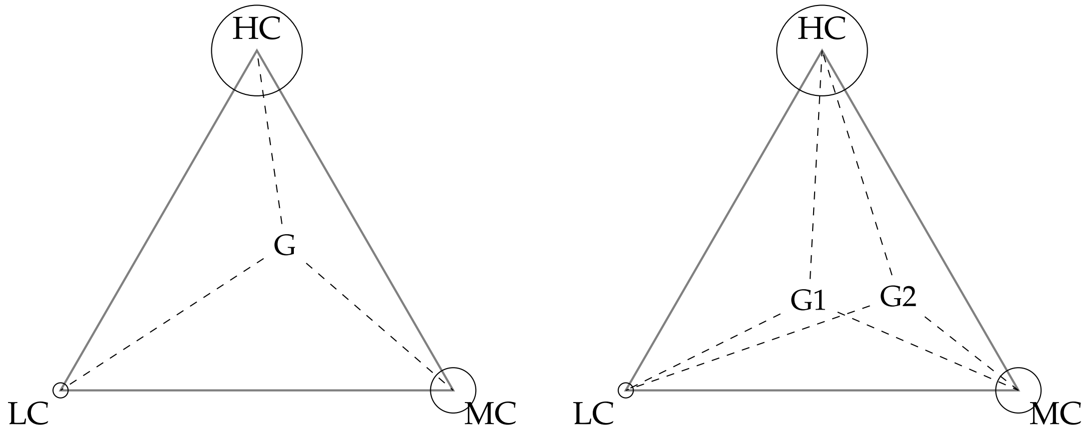
Proposition 1. *Let $\mathcal{J}^*_s$ be the optimal set of active markets. The optimal social location for a firm is a relative size-based weighted average of the social distance between the firm owner and*

[17] The firm size is proportion to z^{σ}, thus this functional form captures the fact that large firms need to hire more employees to move the same distance than the small firms.

[18] For example,firm owner LC with quality z when moving Δx along the second dimension pays $\tilde{\gamma}_{z,s,k} \times \Delta x$ to MC workers, whereas, when moving along the first dimension pays both $0.5\gamma \times \Delta x \times L$ to MC and $0.5\gamma \times \Delta x \times L$ to HC workers.

[19] The fixed component, $f^d_{s',s}$, and variable component, $\Phi(\Delta\mathcal{X}_{z,s};\Gamma_{z,s})$, together determine changes in social location and thus, product appeal. Conversely, we can also assume that an initial certain number of out-group hires ($f^d_{s',s}$) allow firms to capture out-group demand but do not affect the demand from the other groups, that is, *demand spillovers* only exist above certain thresholds.

Figure 1: Optimal Social Location for Firms



(a) No Moving Costs (b) With Moving Costs

Notes. In both figures, we consider a case with three groups (coordinates): LC (1,0,0), MC (0,1,0), and HC (0,0,1); that are represented in a three-dimensional Euclidean space ($\mathcal{R}^3_+$). The size of the circles at the edges represents the relative size of each group. We consider a firm with large z, such that $\underline{X}_{ss'} \to \mathbf{0}$. Figure 1a shows the case where all firms face no cost to locate in group space. All firms choose the same identity represented by the point G. 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 shows the case where the cost of moving in group space is positive. Firm G2 belongs to the firm owner who 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 who was born as LC. Her firm chooses position G1 which is closer to HC and MC consumers to take advantage of their size but not as much as in the case of no moving costs.

the consumer groups and is given by

$$\Delta\mathcal{X}^*_{z,s,k} = \max\left[\underline{X}_{z,s,k}, \frac{\sum_{\mathcal{J}^*_s} \lambda_{s',s,k} d_{s',s,k}}{\sum_{\mathcal{J}^*_s} \lambda_{s',s,k}}\right], \quad \forall \quad k. \tag{12}$$

The above proposition states that firms move closer to the group with the biggest size by hiring more workers from that group. The distance moved decreases with the cost shifter γ. In the baseline framework, as $\gamma = 1$, market size is the sole determinant of the firm's social location $\Delta\mathcal{X}^*_{z,s}$, or in other words, the optimal cross-group hiring. For large enough firms within a group, $\underline{X}_{z,s,s'}$ is small, and therefore, the ethnic workforce composition is the same irrespective of their quality z. In the absence of convex moving costs, $\gamma = 0$, all large enough firms would choose the same ethnic workforce composition – thus there is no cross-group heterogeneity (see Figure 1a). When $\gamma \neq 0$, firms of all sizes will differ in their outcomes – thus there is heterogeneity among large firms from across different groups (see Figure 1b).

Proposition 2. *Under partial equilibrium, the elasticity of firms' revenues earned from the target group s' to income shocks $I_{s'}$ to the target group (all else constant) is given by*

$$\frac{\partial \log r(z,s',s)}{\partial \log I_{s'}} = \underbrace{\frac{\partial \log \kappa_s'}{\partial \log I_{s'}}}_{\textbf{\textit{Size effect}}} + \underbrace{2\tilde{\beta} \sum_{k \in \mathcal{K}} d^*_{z,s',s,k} \frac{\partial \Delta \mathcal{X}_{z,s,k}}{\partial \log I_{s'}}}_{\Delta \textbf{\textit{Optimal Location}}} \tag{13}$$

The derivation of each component is in the appendix. The elasticity of firms' overall revenue is then given by

$$\frac{\partial \log R(z,s)}{\partial \log I_{s'}} = \frac{\partial \log r(z,s',s)}{\partial \log I_{s'}} A_{s',s}, \tag{14}$$

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

In Proposition 2, we provide the expression for the revenue elasticity to an income shock to group s'. There are two effects, the direct effect and the indirect effect. The former comes from the increase in the size of the group, all else constant. The indirect effect captures the fact that firms change their workforce composition, i.e., hiring more employees from the group s' that witnessed a positive shock to income. The pass-through of changes in workforce composition to a firm's revenues depends on $\tilde{\beta}$, the social identity parameter or the trade resistance parameter. This framework nests the standard trade model, where the micro-trade elasticity is given by the direct effect and $\frac{\partial \Delta \mathcal{X}_{z,s,k}}{\partial \log I_s'} = 0$.

The elasticity of the firm's total revenue to an income shock to group s' is a product of the elasticity of revenues from that group and the share of the firm's revenue attributed to that group. This implies that firms that sell relatively more to the group that experiences an income shock have a higher elasticity of overall revenues. Next, we are interested in how these elasticities respond to changes in the preference for the trade resistance parameter $\tilde{\beta}$.

Proposition 3. *Under the assumption that $\frac{\partial \Delta \mathcal{X}_{z,s,k}}{\partial \tilde{\beta}} \approx 0$, the elasticity of revenue shares to the trade resistance parameter $\tilde{\beta}$ (with constant σ) is given by*

$$\frac{\partial A_{s',s}}{\partial \tilde{\beta}} = A_{s',s} \sum_{s''} A_{s',s''} \left((d^*_{s',s''})^2 - (d^*_{s',s})^2 \right) \tag{15}$$

Corollary 1. *If the cost of changing a firm's social identity is high, the share of revenues coming from firms' own group consumers is increasing in $\tilde{\beta}$. Thus, the elasticity of the in-group firms' total revenues with respect to an income shock to group s is higher than the elasticity to an income shock with respect to any other group.*

To summarise, we show that, under certain conditions, the total revenue elasticity of a firm to an income shock to group s is high for firms that, ex ante, have high revenue shares from group s. Furthermore, this difference increases with the taste for social identity $\tilde{\beta}$. We will use these predictions to estimate $\tilde{\beta}$ in the next sections.

3.3.2 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 labour. 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

$$\int_z \max\left[0, \frac{\pi(z,s)}{\delta}\right] = w_s f_e, \quad \forall \quad s \in \mathcal{S}. \tag{16}$$

In equilibrium, exiting firms are replaced by new entrants such that the firm distribution remains stationary. We denote by $\{L_s^e\}_{s\in\mathcal{S}}$ the mass of workers used for entry and by $\{M_s^e\}_{s\in\mathcal{S}}$ the mass of potential entrants each period. Let us denote the probability of survival of entrants by ω_s^e. In a steady state, the mass of operating firms, M_s, remains constant for each group, such that $\delta M_s = \omega_s^e M_s^e$.

3.4 Equilibrium

Given the exogenous quality distribution $G(z)$, that is same across groups, the equilibrium in this economy is a set of prices that includes wages for labour $\{w_s\}_{s\in S}$, prices for each variety of differentiated good $\{p(z,s,s')\}_{s\in S}$, and price for homogeneous good P_H, and a set of quantities that includes consumption quantities $(C_H, \{c(z(\omega),s,s')\}_{s\in S})$, output quantities $(y_H, \{\ell(z,s,s'), y(z,s,s')\}_{s\in S})$, and mass of active firms $\{M_s\}_{s\in S}$, mass of entrants $\{M_s^e\}_{s\in S}$, and labour used for entry $\{L_s^e\}_{s\in S}$, such that households maximize utility according to (1), producers of differentiated good varieties maximize profits and charge the constant markup price, product markets clear for the homogeneous good and for each of differentiated goods, the free entry condition holds, and labour market and product market clears for each group.

3.5 Taste and Transportation Costs: An extension

In the previous section, we only focused on taste-specific demand shifters. However, this may be confounded by transportation costs if consumers of different groups reside in different locations (segregated regions). The trade across groups will be affected by both tastes and transportation costs. Under certain assumptions, we show that the overall trade barrier for a firm can be summarised as a composite $\Lambda_{s',s} = \tau_{s',s}^{1-\sigma}\Psi_{s',s}^{\sigma}$, where $\tau_{s',s}$ captures the transportation costs. We define a parameter $\hat{\beta} = \tilde{\beta}\sigma + \nu(\sigma - 1)$ that captures the strength of cross-group trade barriers in this setting, where $\tilde{\beta}$ is taste elasticity and ν is the transportation cost elasticity. When $\nu = 0$, we are in the benchmark case. We provide details in Appendix A.

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

By this extension, the demand segmentation may be driven by two forces. Thus, to identify the taste channel, we need to rely on a setting where geographical segregation by group is less of a concern. We do that by exploiting village-level population share by groups in India and by replicating our empirical results in assimilated regions (see Section 5.3 for more details).

3.6 Taking Stock

We have developed a theory linking homophily in consumer demand to firms' incentives to trade across castes, affecting firm creation, equilibrium firm size distribution, and employee group composition. At the heart of our model is the insight that a taste for identity in consumer demand limits firms' market access across castes. Small firms only sell to in-group consumers, while large firms sell to diverse groups and have a diverse workforce.

These trends align qualitatively with established observations about firm dynamics in India and other emerging markets where group identities are strong and persistent, characterised by the prevalence of small, non-expanding businesses, an abundance of subsistence producers, and homophily in employee hiring. In the following sections, we will assess the extent to which this mechanism can provide a quantitative explanation for the disparities in firm size in India. We start by providing details on institutional background of our empirical setting.

4 Data and Measurement

We integrate data from various sources to empirically examine the relationship between firm size, employee caste composition, and local demand distribution. We start by summarizing the firm-level data.

4.1 Micro Small and Medium Enterprise (MSME) Data

The MSME dataset is a nationally representative sample of micro, small, and medium-sized firms in India for 2006-07, offering the unique advantage of including the caste of both the enterprise owner and employees—missing in other popular datasets like ASI and Prowess. Unlike ASI and Prowess, which focus on large firms, MSME excludes enterprises above a certain capital threshold. This is less of a concern for us, as large firms, typically multi-establishment, are less likely to suffer from identity bias compared to mostly single-establishment firms in our sample.[20] Also, large firms represent a relatively small fraction of total firms in India (Hsieh and Olken, 2014).

The MSME survey consists of registered and unregistered firms. This classification is defined under the Factories Act 1948. We focus on registered firms as they are larger and less likely to be subsistence enterprises. The dataset provides the caste and gender of the firm owner and employees, along with balance-sheet variables. We keep firms in rural areas and drop sectors that include the manufacturing of food and food products. After cleaning, we have 349,715 firms employing approximately 3 million employees

[20] In the model, we also include a homogeneous goods sector, which is not subject to consumer bias.

in 178 sectors producing 4,110 distinct products and services; see Appendix B.1.1 for more details. We provide the distribution of employees and revenues in Table 1. On average, LC-owned firms are small; however, there is spatial variation. Figure 3a plots the difference in revenues between HC- and LC-owned firms, highlighting substantial variation 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 any particular castes, districts, or sectors (see Figure **??** in appendix), while there is substantial heterogeneity. We find that homophily in hiring declines faster among LC-owned firms and in customer-facing or contact-intensive sectors (see Table B.2 in appendix).

Balanced panel 2004-2006: 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 temporal and geographical variation in demand and revenues across castes.

Prices and quantities: The MSME data provides firm-level information on revenues and quantities for four main *products* and three main *input materials* for the cross-section of firms during 2006-07. We compute average product prices by dividing product revenues by quantities. Figure B.6 provides the distribution of the raw and residualized prices. 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 effect and quantity effect of demand shocks.

4.2 Consumption Data

We use data from the National Sample Survey (NSS) Household Consumer Expenditure survey on households, their consumption, and their demographics. We use the *monthly per capita expenditure* (MPCE) in seven categories - All, Food, Non-food, Fuel & light, Clothing, Footwear, and Durables. To obtain real consumption, we divide nominal consumption by the state-level "Consumer Price Index for Agricultural

All firms	Mean	Median	p5	p95	N
Emp. All	5.4	2	1	15	349715
Emp. LC	1.2	0	0	4	349715
Emp. MC	1.9	1	0	5	349715
Emp. HC	2.3	0	0	7	349715
Emp. Own-C (%)	75	100	0	100	349710
Revenue (10^4)	327	13.0	2.9	456	349715
Materials (10^4)	210	4.1	0.28	266	349715

Table 1: Firm-size distribution

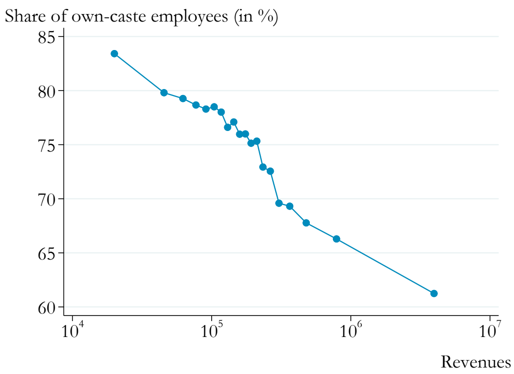


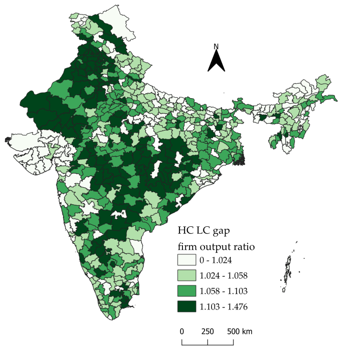
Figure 2: Own-Caste employee share

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 count LC, MC and HC employees within a firm respectively. 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 sector fixed effects. Sampling multipliers are applied.

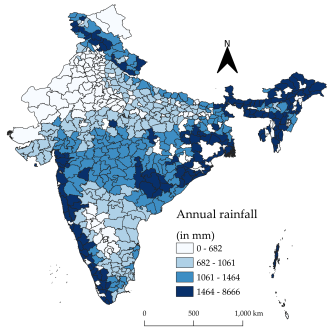
Labourers." The descriptive statistics are provided in the appendix in Table B.4. On average, HC households' monthly per-capita expenditure (1234.5 Indian rupees) is nearly double that of LC households (627.1 Indian rupees).

Identity Engel Curves: We measure the sensitivity of consumer expenditure on a product to the perceived social distance. We define a product's social distance by computing the caste employee share in each product category by district using the MSME data, resulting in three measures of perceived social distance, each ranging between 0 and 1. The first measure is the ratio of the share of LC employees to the combined share of LC and HC employees in each product market. We hypothesise that as this ratio increases, the expenditure of LC consumers on that product declines relative to HC consumers. The results, presented in Table B.5 in the appendix, show that as relative HC employee share increases, the LC consumer's consumption declines at a rate of -0.126 relative to HC consumers. We repeat this exercise for different measures of social distance and find that the slope lies between -0.126 and -0.156. This provides motivating evidence that product demand is influenced by perceived identity. In section 6.3, we discuss how these estimates relate to our eventual estimation of taste elasticity β.

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



(a) Output-difference between HC- and LC-owned firms



(b) Annual 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 annual rainfall received by Indian districts, in millimeters of rainfall.

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 millimetres of annual rainfall, for the year 2006-07. The plot shows a lot of variation across districts in India. For the analysis, we follow Jayachandran (2006) and define rain shock as equal to +1 for positive shock, -1 for negative shock, and 0 otherwise.[21]

5 Empirical Strategy

Motivated by the model predictions documented in Section 3, 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.[22] 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

[21] See Appendix for more details.

[22] A district is an administrative unit in India analogous to counties in the US system. Approximately, there were an average of 17 districts per state, with an average of 1.75 million total residents per district.

different caste groups across different sectors.

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.6 for evidence on wages). LC households predominantly engage in agricultural labour. To do this, we estimate the following equation:

$$log(c_{ht}) = \alpha + \beta_1 \cdot Rainshock_{dt} + \beta_{2,i} \cdot Rainshock_{dt} \times caste_i + \delta_{d,t} + \delta_i + \epsilon_{ht}, \quad (17)$$

where h denotes household, d denotes the district, and t denotes year. The regression includes caste, and district $\times$ year fixed effects to control for any time-varying district-specific characteristics. 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 rainfall shock across districts, to obtain a plausibly causal interpretation. 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-owned firms, we check whether the above shift in LC households' demand particularly makes LC-owned firms bigger. To do so, we estimate the following equation:

$$log(y_{ft}) = \alpha + \beta_1 \cdot Rainshock_{dt} + \beta_{2,i} \cdot Rainshock_{dt} \times caste_i + \delta_{d,t} + \delta_{s,t} + \delta_i + \epsilon_i, \quad (18)$$

where f denotes the firm and s denotes the sector or product. We estimate the effect of higher rainfall on firm-level (y): (1) revenue and (2) material input. The regression includes caste, and district $\times$ year fixed effects. The regression also includes a 4-digit sector $\times$ 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-owned 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:

$$log(y_f) \;=\; \alpha + \beta_1 \cdot Rainshock_d + \beta_{2,i} \cdot Rainshock_d \times caste_i + \delta_d + \delta_s + \delta_i + \epsilon_i, \quad (19)$$

where firm-level outcome variables are caste-specific employee shares. The only difference relative to Equation (18) 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-owned firms. Motivated by model predictions, we focus on large firm that are more likely to be involved in cross-group trade and hire employees outside their group. 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.

Note that, using the above fixed effects specification, we cannot 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 HC firms dominate, the incentive for HC firms to hire LC workers is low as LC consumers have low-to-no outside options other than consuming goods produced by HC firms. We define a market as a Cartesian product: district $\times$ product. 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:

$$log(y_f) \;=\; \alpha + \beta_1 \cdot Rainshock_d + \beta_2 \cdot Rainshock_d \times Competitive_{d,s} + \delta_d + \delta_s + \epsilon_i, \quad (20)$$

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. Within the firms owned by HC, we run the regression:

$$log(y_f) \ = \ \alpha + \beta_1 \cdot Rainshock_d + \beta_2 \cdot Rainshock_d \times Contact\text{-}Intensive_s + \delta_d + \delta_s + \epsilon_i. \tag{21}$$

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 2 presents the results from estimating Equation (17), and shows the asymmetric effect of higher rainfall on Monthly Per Capita Expenditure (MPCE). LC households' consumption increases by 8.6% in districts with a positive rainfall shock relative to 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 (see Figure 4). We also find that MC household's demand goes up, but the elasticity is smaller relative to the LC households.[23] Appendix Table B.6 and Appendix Figure B.3 present consumption elasticities, controlling for the household head's meals and land ownership (as proxies for wealth), as well as education interacted with rainfall shocks. Our estimate remains stable, suggesting that the wealth channel is not driving the rainfall-induced demand shocks.

5.2.2 Firm revenue elasticity

The above evidence shows that higher rainfall induces a positive effect on the local economy with a shift in demand for products, largely driven by LC households. We now evaluate the firm outcomes. Table 3 presents the results from estimating Equation (18), and shows the asymmetric effect of higher rainfall on firms' outcomes. We document an increase in the revenue of LC-owned firms by 13.4%, relative to HC-owned firms in districts with positive rainfall shock. Consistent with our hypothesis,

[23]In a series of robustness check, we find LC households' consumption response to be much more robust than the MC households,' see Appendix B.1.

	log *MPCE*
Rainshock × MC	0.073***
	(0.022)
Rainshock × LC	0.086***
	(0.027)
Observations	117,772
R-squared	0.326
Controls$_{it}$	✓
Caste FE	✓
District × Year FE	✓

Table 2: Consumption

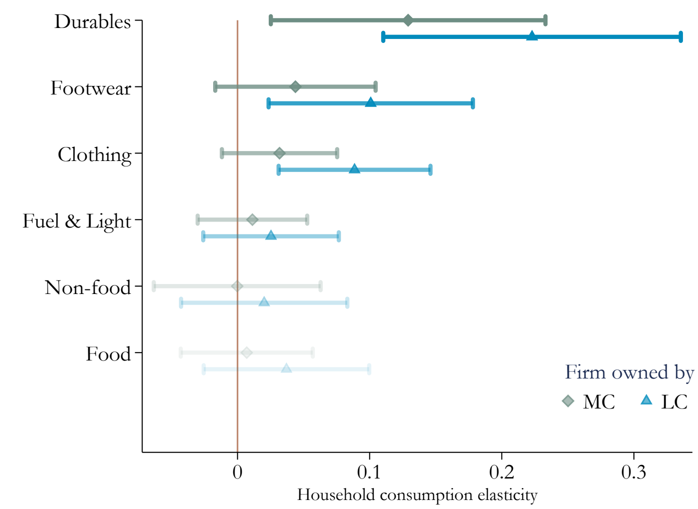


Figure 4: Heterogeneity in consumption

Notes. Table 2 presents the elasticity of monthly per-capita consumption of households in our sample. The regressions are of household-level variables (logarithmic) on rainfall shock by caste. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at the district level in all regressions, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Figure 4 presents the heterogeneity in consumption elasticities with values on the x-axis and the consumption groups labeled on the y-axis. We control for household-level wealth (using meals per day and land owned) and education, and their interactions with rainfall. We also control for caste, district, and year fixed effects.

this observation suggests that LC-owned firms gain the most due to a positive shift in demand from LC households. The LC-owned firms also witness a 20% rise in material input purchase, relative to HC-owned firms. These results remain robust to the inclusion of product × year fixed effects. Focusing on product markets where LC household consumption response is highest, in Columns 5 to 8, we find that LC-owned firms have substantially higher revenue elasticity (+ 5 percentage points relative to the baseline effect).

5.2.3 Workforce Composition

To further shed light on the increase in the size of LC-owned 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. Guided by theory, first, we run

Table 3: Firm-size Elasticity in rural India

	All Sectors				LC's High-consumption Sectors			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	Revenues	Inputs	Revenues	Inputs	Revenues	Inputs	Revenues	Inputs
Rainshock × MC	0.116***	0.162***	0.103***	0.138***	0.094**	0.121*	0.092**	0.115**
	(0.040)	(0.060)	(0.032)	(0.051)	(0.046)	(0.063)	(0.041)	(0.058)
Rainshock × LC	0.134***	0.200***	0.121***	0.176***	0.184***	0.268***	0.155***	0.227***
	(0.047)	(0.065)	(0.036)	(0.052)	(0.067)	(0.085)	(0.049)	(0.066)
Observations	950,345	941,873	947,614	939,134	407,531	403,971	406,517	402,952
R-squared	0.512	0.544	0.594	0.610	0.463	0.471	0.543	0.539
Caste FE	✓	✓	✓	✓	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Sector × Year	✓	✓			✓	✓		
Product × Year FE			✓	✓			✓	✓

Notes. The regressions are of firm-level variables (logarithmic) on rainfall shock by caste. Columns 1 to 4 include observations from all sectors. Columns 5 to 8 include observations from the sectors where LC households display high elasticities following the rainfall shock. We control for caste, district × year, and sector × year or product × year fixed effects. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at the district level in all regressions, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

the regressions in Equation (19) among the large firms that hire HC employees and provide results in Table 4. We employ this sample because we want to focus on the changes in the intensive margin of hiring HC workers. The table shows that LC-owned firms are larger in districts with positive rainfall shock, that they have a lower share of HC workers, and that they have a higher share of LC workers (i.e., the caste group that experienced a positive demand shock). All these changes are measured relative to HC firms. This evidence supports the view that when LC consumer demand is high, then the incentive to hire LC workers by large LC-owned firms is also high. In line with our previous results, these changes are much larger in magnitudes in product markets where LC household consumption response is high.

This change in workforce composition is in contrast to the cross-sectional descriptive evidence provided in Section 4.1, where large LC-owned firms are less homophilic and hire more out-group employees. Thus, it suggests that the changes in employee composition are demand-driven rather than supply-driven. We will discuss this in detail in the next section.

Workforce composition of HC firms in Competitive markets. We run regressions using Equation (20) specifically within the HC firms. Market delineations were defined by district × product. Figure 5 presents the findings. The results show that in

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

	All Sectors				LC's High-consumption Sectors			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	Total Employees	HC Share	MC Share	LC Share	Total Employees	HC Share	MC Share	LC Share
Rainshock × MC	-0.019	0.008	0.006	-0.013	-0.029	0.009	0.005	-0.014
	(0.034)	(0.015)	(0.011)	(0.008)	(0.042)	(0.018)	(0.013)	(0.009)
Rainshock × LC	0.081***	-0.045**	-0.010	0.055***	0.127***	-0.034*	-0.007	0.041***
	(0.030)	(0.021)	(0.010)	(0.017)	(0.041)	(0.018)	(0.012)	(0.015)
Observations	162,719	162,719	162,719	162,719	72,979	72,979	72,979	72,979
R-squared	0.439	0.300	0.261	0.228	0.410	0.323	0.274	0.253
Caste FE	✓	✓	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓	✓	✓	✓

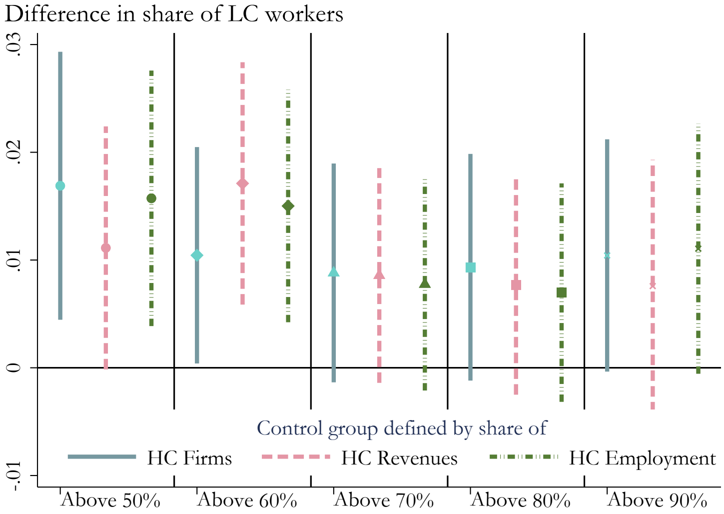
Notes. The regressions are of using Equation (19) among the firms that hire HC employees. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

competitive markets, HC-owned firms hire more LC employees when LC household demand is high, compared to HC-dominated markets. This indicates that competition drives demand-led diversification, as firms capture consumer demand via targeted hiring.

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

Workforce composition in contact-intensive sectors. Next, we highlight how the changes in the workforce vary across sectors. Appendix Table B.3 describes the labour share of workers in our sample of firms across different sectors. In contact-intensive sectors, we expect castes with less appeal to the wealthier consumers to have a lower labour share than in sectors that are less contact-intensive, as firms try to project themselves closer to the wealthier castes. Consistent with this, Columns 1-2 show that LC labour share is higher in less contact-intensive sectors while HC labour share is higher in more contact-intensive sectors. On the other hand, when it comes to sectors where LC households display high elasticities following the rainfall shock, we observe the opposite pattern. LC labour share is higher in sectors where LC consumers show a higher consumption than in other sectors, while HC labour share is substantially lower in sectors where LC consumers show a higher consumption than in other sectors.

Figure 5: LC Employee Share Among HC-owned Firms in Competitive Markets



Notes. The figure presents the β_2 (y-axis) from regressions using Equation (20) among the HC firms. The dependent variable is LC employee share. We defined the market by district $\times$ Product. 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 line). The band around the point estimate represents the 90% confidence interval.

In our final strand of evidence, we run the regressions using Equation (21) within HC-owned firms and provide results in Table 5. The results show that HC-owned firms hire more LC employees in customer-facing or contact-intensive sectors in regions where LC household demand is high. This is in contrast to the baseline cross-sectional evidence provided above (see Appendix Table B.3), where LC employee share in contact-intensive sectors is lower on average. This again suggests the demand channel at play. The key takeaway 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 household consumers. We use two different definitions of contact-intensive industries (list is provided in Appendix B.1) and results remain robust. In line with our previous results, these elasticities are much larger in product markets where LC household consumption response is high. All three pieces of evidence taken together highlight that firms capture diverse consumer markets by hiring employees from the target consumer caste.

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

	All Sectors		LC's High-consumption Sectors	
Share of LC workers	(1)	(2)	(3)	(4)
Rainshock × Contact-intensive 1	0.013**		0.024***	
	(0.006)		(0.008)	
Rainshock × Contact-intensive 2		0.014**		0.024***
		(0.007)		(0.008)
Observations	131,353	131,353	58,114	58,114
R-squared	0.203	0.203	0.241	0.241
District FE	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓

Notes. The regressions are of using Equation (21) among HC-owned firms. The list of industries in Contact-Intensive sectors is provided in Appendix B.1. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5.3 Discussion on Empirical Results and Robustness Checks

In this subsection, to address concerns regarding the interpretation of our main results, we discuss alternative mechanisms and check for evidence on them.

Spatial Segregation. Is demand segmentation just represent the underlying spatial segregation? The observed segmentation could be attributed to two underlying mechanisms: (a) homophily in consumption, as hypothesized, 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). We carry out a supplementary analysis by excluding observations from districts that exhibit high levels of geographical segregation.[24] The results from 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.7).

Price and Quality. Is the increase in LC revenue due to rainfall shocks driven by changes in prices or product quality? Using data on product prices, we find no significant impact of rainfall on output prices in our sample. Additionally, using input prices as a proxy for product quality, we observe no significant effect of rainfall on

[24]Our measure of geographical segregation quantifies the standard deviation of the LC population share within a district (Figure B.4 in the appendix shows the distribution of our measure across districts in Tamil Nadu).

quality either (see Table B.8).

Information Frictions. We explore the role of such information frictions in explaining homophily in consumption: when it is difficult to assess product quality, consumers may use caste as a proxy. To test this, we use two proxies: (i) Rauch (1999) measure of product differentiation and (ii) price dispersion within a narrowly defined product market. According to such a hypothesis, in less differentiated products, where information frictions are minimal, LC-owned firms' revenue elasticity should be similar to that of HC-owned firms. However, our results show that LC-owned firms' revenue elasticity is significantly higher than HC firms', and homophily in consumption remains strong, even in less differentiated products. (see Tables B.10 and **??**).

Financial Frictions. Do rainfall-induced income shocks alleviate credit constraints for households and, thus, stimulate firm growth? We investigate this by examining the heterogeneous effects across firm size. We find larger effects among relatively large firms, in contrast to the literature on financial frictions, which suggests small firms are more likely to be constrained. Additionally, we find no significant effect of rainfall neither on formal nor on informal loans. Finally, we compute the marginal revenue product of capital (MRPK), as constrained firms are expected to have a high MRPK, according to the literature on misallocation (Hsieh and Klenow, 2009). If financial constraints are relaxed by rainfall shocks, MRPK should decline. However, we find no significant effect of rainfall on MRPK (see Table B.8).

Wealth-based Demand Channel. Can segregation by wealth explain our results – LC households, being poorer, consume low quality products that are mostly produced by LC-owned firms? We have already shown that LC demand response is not driven by wealth status or education (see Appendix Table B.6 and Figure B.3). Next, to parse out how rainfall effects differ by LCs' industry of specialization, we differentiate between industries concentrated with LC-owned firms and otherwise and find no significant differences in the results between these two industry segments (see Appendix Table B.9). Finally, to capture the product quality dimension, we consider the set of markets where prices of inputs used by LC- and HC-owned firms in production are similar (product quality being proxied with input prices as described above). We do not find that the revenue effect among LC-owned firms is lower than our baseline results, thus weakening the evidence for a wealth-based demand channel (see Appendix

Table B.12).

Export-led Demand Shocks and Firm Outcomes. Do expanding firms, regardless of the origin of demand shocks, consistently hire LC workers because they are relatively less expensive than HC or MC employees? To test this hypothesis, we carry out a placebo exercise by exploiting foreign product demand shocks, which are caste-neutral by nature. Using data on export values provided in the MSME survey, we examine whether firms' exports and revenues respond positively to exogenous changes in foreign country demand. Our findings show that foreign demand shocks are indeed expansionary; however, they do not lead to an increased demand for LC employees. The LC employee share remains constant across HC-owned firms. This suggests that unlike rainfall-induced demand shocks, firms do not adjust their employee mix when responding to increases in foreign demand.

Alternative effects of rainfall. One concern might be that rainfall shocks affect the non-agricultural labour market through alternative mechanisms. For instance, increased rainfall boosts demand for labour in the agricultural sector, which aligns with the observed rise in agricultural wages. This could initially attract lower caste (LC) workers into agricultural work, where they have historically comprised a larger share of the labour force. As a result, non-agricultural firms may experience a reduced supply of LC workers, potentially leading to a decrease in LC labour within these firms. However, contrary to this expectation, we observe that rainfall positively influences LC hiring by non-agricultural firms, suggesting that this mechanism does not explain our observations. In fact, the asymmetric rise in LC employment in non-agricultural sectors supports a demand-driven hiring explanation.

Additionally, rainfall-induced wage increases in agriculture might spill over into other sectors, enhancing workers' bargaining power and providing a more attractive outside option. This scenario would imply an asymmetric rise in wages for LC workers in non-agricultural sectors and potentially a decline in demand for these workers due to their improved wage bargaining power. However, one may not see a rise in the non-agricultural sectors' wages if there is enough slack labour as highlighted by Breza et al. (2021). While an increase in LC worker's wages in non-agricultural sectors relative to other workers may decrease their demand; on the contrary, we observe that non-agricultural firms hire more LC workers. This further supports a demand-driven hiring mechanism is driving our results.

Alternatively, rainfall may expand the size of all markets due to an unobserved factor. Yet, this would not explain the specific increase in LC employment relative to other caste groups. Even if rainfall enlarges LC-intensive markets specifically, our analysis shows that the effect persists across sectors where LC firms are not the primary players, reinforcing the importance of demand-driven hiring mechanism in explaining our findings.

6 Model Quantification

In this section, we use our framework presented in Section 3 and the empirical estimates from Section 5.2 to quantify the effect of taste for identity in demand on cross-caste trade, 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 elabourating 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 normalised 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 counterparts in the model. Next, we illustrate the fixed parametric values.

6.2 Fixed Parameters

In the model, we had allowed for heterogeneity across castes in 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 labour from their caste for production. The productivity of the homogeneous good sector A_H is normalised 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 normalised to one, following

Bai et al. (2019). Further, following the previous literature (see, for instance, 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 labour 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 that are the same across castes; probability of firm death δ, fixed cost of operating $c_f = f^d_{s,s}$, 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 $c_x = f^d_{s,s'}$. We estimate c_x in terms of the fixed production costs $c_x = xc_f$.

Further, labour 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. We will assume that the revenue increase among LC-owned firms is uniformly spread across all months, which gives us a 1.11% monthly increase in revenues. In the model, we match this revenue increase in partial equilibrium, i.e., all else equal, an 8.6% percent increase in the market demand from LC consumers and a 7% increase in the market demand from MC consumers should increase LC-owned firms' revenue by 1.11%. The model observes a monotonic relationship between the taste for an identity β and revenue elasticity (see Figure 6a).

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 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 6b), 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 6d).

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 sell 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 et al. (2022).[25] 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 6c).

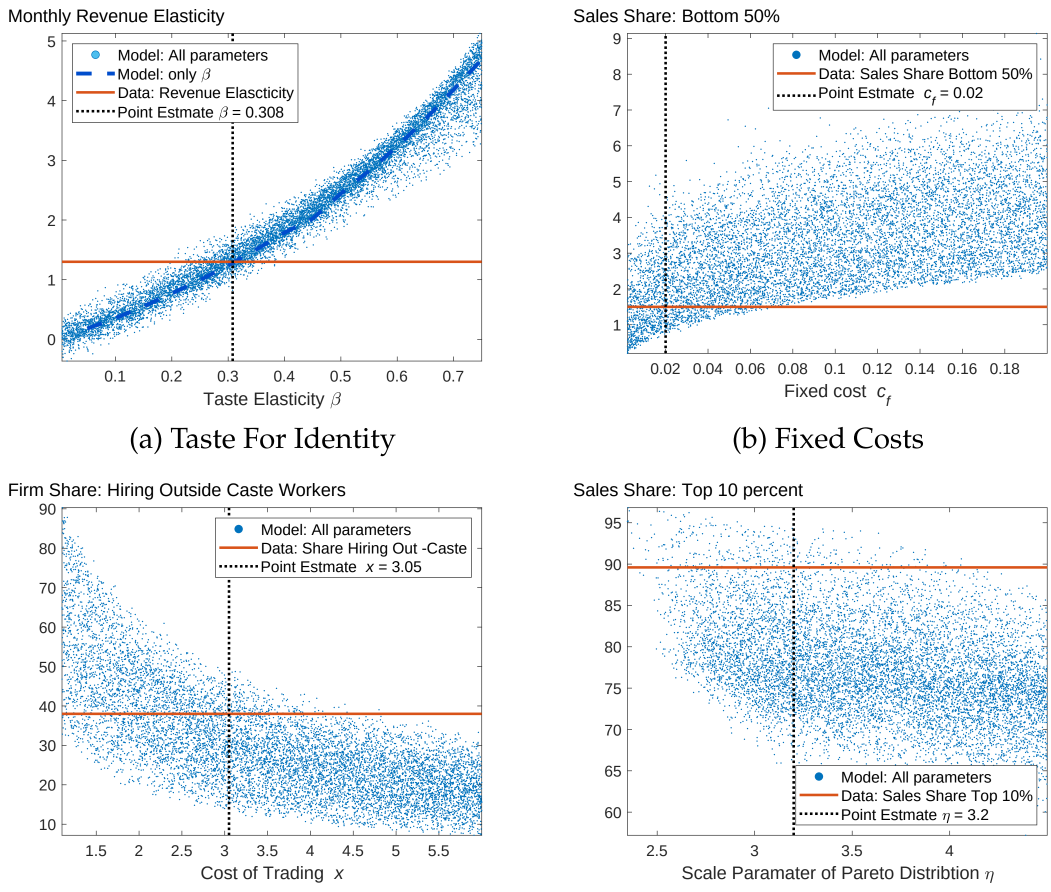
Table 6: Fitted Model Parameters and Targetted Moments

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 dist.	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					
$\{L_{LC}, L_{MC}, L_{HC}\}$	{0.30,0.34,0.36}	Household Labour Endowment			
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 13.4% (the yearly relative revenue increase for LC-owned firms as in Table 3) by 12 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 population 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).

[25] Their dataset is based on firm-to-firm trades within West Bengal. The difference in probability of trading between the studies may be attributed to the fact that our data is skewed towards small firms; while their data is skewed towards large firms.

Figure 6: 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 tessaract $[0.005, 0.15] \times [2.0, 4.5] \times [0.002, 0.20] \times [1.1, 6.0]$. 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.

6.3 Results: Taste for Identity, Firm Size and Aggregate Economy

The estimated parameters and the corresponding moments are provided in Table 6. The value of taste for the identity parameter is estimated to be 0.308, the fixed cost of production is 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 to be 3.20.

The model finds a substantial taste for identity in the economy with demand decaying at a rate of 30.8% over the social distance between castes. In Figure 7a, we compare the demand decay rate for different levels of taste for identity parameter β and show that the relationship between demand decay and taste for identity is convex. The estimate of β is substantially higher than our OLS estimates of Engel curves in Section 4.2. This highlights that the OLS estimates were downward biased, emphasising the usefulness of demand shocks.[26]

Within the benchmark economy, the distribution of firms among different castes exhibits a magnitude that exceeds the proportionality to their respective labour endowments. This finding draws parallels with the work of Helpman et al. (2004) in the field of exports. Consequently, HC firms assert control over the majority of the market share within the differentiated goods sector, whereas LC firms specialise 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 labour endowment. This preference for identity acts as an implicit trade friction, resulting in a reduction of real income for all caste groups.

In Figure 7b, we analyse the aggregate implications of homophily in consumption. Initially, this discourages firms from venturing into markets that are socially distant, leading to a small firm size, and markets remain predominantly socially local. This stems from two primary factors: firstly, firms incur a fixed cost when hiring or accessing a socially 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. Figure 7b illustrates this 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, for instance, Melitz, 2003). In this context, when firms hire workers from other castes or sell to other castes, the demand from their group decreases. This trade-off diminishes the incentive for diversification. Both the extensive margin of sales and the intensive margin of improving taste contribute to smaller firm sizes. This outcome provides a demand-driven explanation for the observed phenomenon of small firm size in India (Hsieh and Klenow,

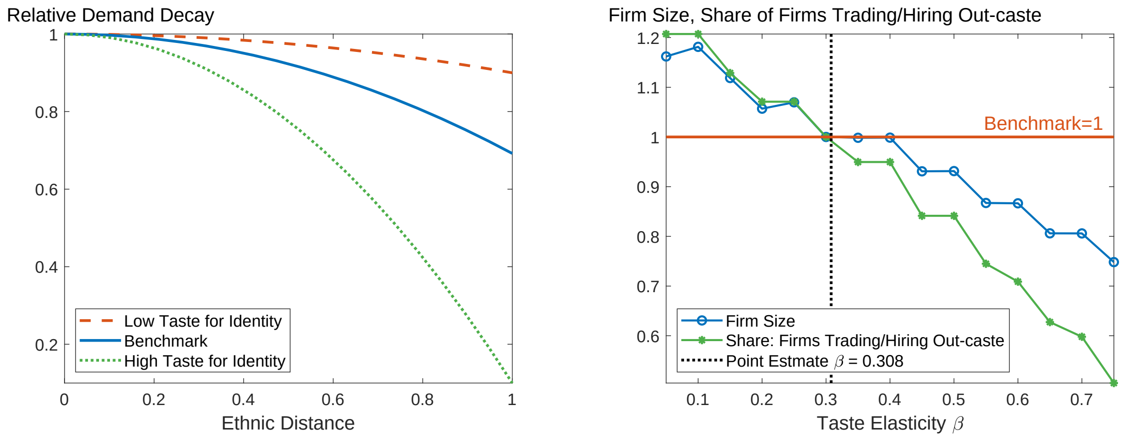
[26] The downward bias in the OLS can be driven by omitted variables, such as quality, that are likely to be negatively correlated with perceived social distance.

2014). We provide corroborative empirical evidence for this mechanism. In Table B.17 in the appendix, we find that lower heterogeneity in the caste composition of a region's population is associated with larger firms, and this correlation is stronger in the customer-facing or contact-intensive sectors.

Given our calibration, we can now compare how much small and large firms differ in terms of cross-caste hiring. By assumption, small firms – that only sell to the in-group consumer – do not hire out-caste workers, thus the share of own-caste workers is 1. Among large firms, the share of cross-caste employees is 4.5 percent. Thus, the model suggests that the demand channel can explain around 10-15% of the workforce diversification among large firms.

Finally, we find that preferences for in-group firms significantly affect aggregate income. We measure real income for each caste group and then compute the population-weighted average. Using a set of counterfactual experiments, we document that doubling the homophily in consumption substantially increases trade barriers and decreases aggregate income by around 5%. Although these preferences limit income in our context, they may provide benefits in other situations. Thus, policies designed to address these barriers should be carefully assessed to ensure that their costs do not exceed the potential benefits.

Figure 7: Taste for Identity, Homophily in Hiring and Firm Size



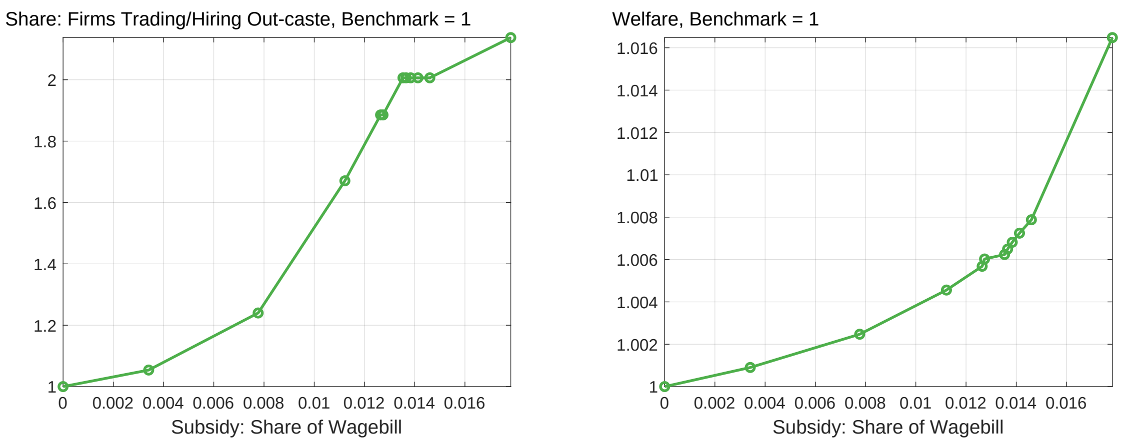
(a) Demand Decay and Taste for Identity (b) Firm Size and Trading/Hiring Decision

Notes: This figure plots the counterfactual aggregate outcomes under different values of parameters β that measure taste of identity in demand. Figure 7a plots the demand decay over the ethnic distance between the firm owner and consumer and Figure 7b 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 normalised to one.

6.4 Cost of Hiring Cross-Caste

Utilising our calibrated model, we ask: Does a reduction in the cost of cross-caste hiring contribute to overall welfare improvement, all else constant? To answer this question, we solve the model for different levels of cost of cross-caste hiring, which capture both the convex cost and fixed cost of cross-caste hiring, respectively. We formally define the cost reduction as a subsidy, that is, 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. Across our policy experiments, the subsidy varies between 0 and 2 percent of the total wagebill of the differentiated good sector.

Figure 8: 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 8a plots the share of firms trading / hiring outside their caste, and Figure 8b computes the real income relative to the benchmark calibration. The benchmark values are normalised to one.

Reduced costs act as an incentive for firms to engage in selling to ethnically distant castes. As this expansion entails hiring more from other castes, homophilly in hiring goes down at the extensive margin (see Figure 8a). 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 labour increases. This results in an increase in the number of firms serving each ethnic market. Consequently, the average firm size increases. This leads to an improvement in welfare, measured by real income for all groups (see Figure 8b).

However, the question remains: Can policies alter the cost of cross-group hiring and promote trade? Recent research indicates that inter-group contact can foster positive social preferences and reduce existing biases. Therefore, short-term subsidies aimed at promoting diversity in the workforce may be desirable.

7 Conclusion

This paper demonstrates that consumer preferences for in-group sellers or products can constrain firms' market access and limit their ability to achieve scale economies; thus influencing firm size distribution and aggregate income. We present a model where firms can sell goods to diverse markets, but it is more costly to sell to socially distant markets where product appeal is lower. Firms, however, hire workers from the targeted consumer group to expand their market access. We show that the optimal hiring of out-group employees is proportional to the market size of each group. Thus, our paper highlights a novel channel of firm expansion and the demand-side determinants of within-firm employee composition.

We exploit a unique data on employer-employee linkages by caste and exogenous shifts in local demand to establish causality. Our analysis reveals that caste-specific demand shocks disproportionately benefit in-caste firms, expanding their revenues relative to others. These demand shocks lead to an increase in homophily in hiring practices within these firms. Additionally, firms owned by other castes respond by hiring more employees from the caste experiencing heightened demand – a reduction in homophily in hiring. Using these micro-elasticities, we calibrate the model, highlighting the significant taste for identity in consumer demand. This impedes firms' ability to trade with a diverse set of consumers and expand. Finally, in a set of counterfactual experiments, we show that reducing the cost of cross-caste hiring fosters cross-caste trade and enhances consumer welfare.

References

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

AFRIDI, F., A. DHILLON, AND S. SHARMA (2020): "The ties that bind us: social networks and productivity in the factory," .

AGTE, P. AND A. BERNHARDT (2023): "The Economics of Caste Norms: Purity, Status, and Women's Work in India," .

AKER, J. C., M. W. KLEIN, S. A. O'CONNELL, AND M. YANG (2014): "Borders, ethnicity and trade," *Journal of Development Economics*, 107, 1–16.

ALBORNOZ, F., H. F. C. PARDO, G. CORCOS, AND E. ORNELAS (2012): "Sequential exporting," *Journal of International Economics*, 88, 17–31.

ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): "Public goods and ethnic divisions," *The Quarterly journal of economics*, 114, 1243–1284.

ALESINA, A. AND E. LA FERRARA (2000): "Participation in heterogeneous communities," *The quarterly journal of economics*, 115, 847–904.

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

ALFARO-URENA, A., J. CASTRO-VINCENZI, S. FANELLI, AND E. MORALES (2023): "Firm export dynamics in interdependent markets," Tech. rep., National Bureau of Economic Research.

ALLEN, T. (2014): "Information frictions in trade," *Econometrica*, 82, 2041–2083.

ALSAN, M., O. GARRICK, AND G. GRAZIANI (2019): "Does diversity matter for health? Experimental evidence from Oakland," *American Economic Review*, 109, 4071–4111.

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

ASHER, S. AND P. NOVOSAD (2019): "Socioeconomic High-resolution Rural-Urban Geographic Dataset for India (SHRUG)," .

ASTURIAS, J., M. GARCÍA-SANTANA, AND R. RAMOS (2019): "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 17, 1881–1940.

ATKIN, D., E. COLSON-SIHRA, AND M. SHAYO (2021): "How do we choose our identity? a revealed preference approach using food consumption," *Journal of Political Economy*, 129, 1193–1251.

ATKIN, D. AND D. DONALDSON (2015): "Who's getting globalized? The size and implications of intra-national trade costs," Tech. rep., National Bureau of Economic Research.

BAI, Y., K. JIN, AND D. LU (2019): "Misallocation under trade liberalization," Tech. rep., National Bureau of Economic Research.

BANERJEE, A. AND K. MUNSHI (2004): "How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur," *The Review of Economic Studies*, 71, 19–42.

BARROWS, G. AND H. OLLIVIER (2021): "Foreign demand, developing country exports, and CO2 emissions: Firm-level evidence from India," *Journal of Development Economics*, 149, 102587.

BASSI, V., J. H. LEE, A. PETER, T. PORZIO, R. SEN, AND E. TUGUME (2023): "Self-employment within the firm," Tech. rep., National Bureau of Economic Research.

BENTO, P. AND D. RESTUCCIA (2017): "Misallocation, establishment size, and productivity," *American Economic Journal: Macroeconomics*, 9, 267–303.

BERNARD, A. B., E. DHYNE, G. MAGERMAN, K. MANOVA, AND A. MOXNES (2022): "The origins of firm heterogeneity: A production network approach," *Journal of Political Economy*, 130, 1765–1804.

BHAGAVATULA, S., M. BHALLA, M. GOEL, AND B. VISSA (2018): "Cultural homophily in corporate boards and firm," .

BHARATHI, N., D. MALGHAN, S. MISHRA, AND A. RAHMAN (2021): "Fractal urbanism: City size and residential segregation in India," *World Development*, 141, 105397.

BIDNER, C. AND M. ESWARAN (2015): "A gender-based theory of the origin of the caste system of India," *Journal of Development Economics*, 114, 142–158.

BOKEN, J., L. GADENNE, T. NANDI, AND M. SANTAMARIA (2022): "Community networks and trade," *Warwick Economics Research Papers*.

BOLD, T., S. GHISOLFI, F. NSONZI, AND J. SVENSSON (2022): "Market Access and Quality Upgrading: Evidence from Four Field Experiments," *American Economic Review*, 112, 2518–2552.

BREZA, E., S. KAUR, AND Y. SHAMDASANI (2021): "Labor rationing," *American Economic Review*, 111, 3184–3224.

BRINATTI, A. AND N. MORALES (2021): "Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments," .

BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *The Quarterly journal of economics*, 121, 541–585.

CASSAN, G., D. KENISTON, AND T. KLEINEBERG (2021): "A division of laborers: Identity and efficiency in India," Tech. rep., National Bureau of Economic Research.

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

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

COMBES, P.-P., M. LAFOURCADE, AND T. MAYER (2005): “The trade-creating effects of business and social networks: evidence from France,” *Journal of international Economics*, 66, 1–29.

DAVIS, D. R., J. I. DINGEL, J. MONRAS, AND E. MORALES (2019): “How segregated is urban consumption?” *Journal of Political Economy*, 127, 1684–1738.

DESHPANDE, M. S. (2010): “History of the Indian caste system and its impact on India today,” .

DESMET, K., J. F. GOMES, AND D. M. RICO (2023): “Ethnic remoteness reduces the peace dividend from trade access,” Tech. rep., National Bureau of Economic Research.

DOLEAC, J. L. AND L. C. STEIN (2013): “The visible hand: Race and online market outcomes,” *The Economic Journal*, 123, F469–F492.

DONALDSON, D. (2015): “The gains from market integration,” *Annual Review of Economics*, 7, 619–647.

EINAV, L., P. J. KLENOW, J. D. LEVIN, AND R. MURCIANO-GOROFF (2021): “Customers and retail growth,” Tech. rep., National Bureau of Economic Research.

EMERICK, K. (2018): “Trading frictions in Indian village economies,” *Journal of Development Economics*, 132, 32–56.

ESLAVA, M., J. HALTIWANGER, AND N. URDANETA (2024): “The size and life-cycle growth of plants: The role of productivity, demand, and wedges,” *Review of Economic Studies*, 91, 259–300.

FELBERMAYR, G. J. AND F. TOUBAL (2010): “Cultural proximity and trade,” *European Economic Review*, 54, 279–293.

FISMAN, R., D. PARAVISINI, AND V. VIG (2017): “Cultural proximity and loan outcomes,” *American Economic Review*, 107, 457–492.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): “Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?” *American Economic Review*, 98, 394–425.

——— (2016): “The slow growth of new plants: Learning about demand?” *Economica*, 83, 91–129.

FUJIY, B. C., G. KHANNA, AND H. TOMA (2022): “Cultural Proximity and Production Networks,” in *Research Seminar in International Economics, University of Michigan Working Papers*, 686.

GHOSH, A. (2024): “Religious divisions and production technology: Experimental evidence from India,” *Available at SSRN 4188354*.

GIL, R. AND J. MARION (2022): “Why did firms practice segregation? Evidence from movie theaters during Jim Crow,” *The Journal of Law and Economics*, 65, 635–663.

GORAYA, S. S. (2023): "How Does Caste Affect Entrepreneurship? Birth versus Worth," *Journal of Monetary Economics*.

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

HARDY, M. AND G. KAGY (2020): "It's getting crowded in here: experimental evidence of demand constraints in the gender profit gap," *The Economic Journal*, 130, 2272–2290.

HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2004): "Export versus FDI with heterogeneous firms," *American economic review*, 94, 300–316.

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

HJORT, J., C. SONG, AND C. YENKEY (2021): "Ethnic investing and the value of firms," .

HNATKOVSKA, V., A. LAHIRI, AND S. PAUL (2012): "Castes and labor mobility," *American Economic Journal: Applied Economics*, 4, 274–307.

HNATKOVSKA, V., A. LAHIRI, AND S. B. PAUL (2013): "Breaking the caste barrier: Intergenerational mobility in India," *Journal of human resources*, 48, 435–473.

HOTELLING, H. (1929): "Stability in competition," *The economic journal*, 39, 41–57.

HOTTMAN, C. J., S. J. REDDING, AND D. E. WEINSTEIN (2016): "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 131, 1291–1364.

HSIEH, C.-T. AND P. J. KLENOW (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

——— (2014): "The life cycle of plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084.

HSIEH, C.-T. AND B. A. OLKEN (2014): "The missing "missing middle"," *Journal of Economic Perspectives*, 28, 89–108.

HUNT, V., S. PRINCE, S. DIXON-FYLE, AND L. YEE (2018): "Delivering through diversity," *McKinsey & Company*, 231, 1–39.

JAYACHANDRAN, S. (2006): "Selling labor low: Wage responses to productivity shocks in developing countries," *Journal of political Economy*, 114, 538–575.

JENSEN, R. AND N. H. MILLER (2018): "Market integration, demand, and the growth of firms: Evidence from a natural experiment in India," *American Economic Review*, 108, 3583–3625.

KAUR, S. (2019): "Nominal wage rigidity in village labor markets," *American Economic Review*, 109, 3585–3616.

KELLEY, E. M., G. V. LANE, M. PECENCO, AND E. A. RUBIN (2024): "Customer Discrimination in the Workplace: Evidence from Online Sales," Tech. rep., National Bureau of Economic Research.

LOWE, M. (2021): "Types of contact: A field experiment on collaborative and adversarial caste integration," *American Economic Review*, 111, 1807–1844.

MADHESWARAN, S. AND P. ATTEWELL (2007): "Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey," *Economic and political Weekly*, 4146–4153.

MCCALLUM, J. (1995): "National borders matter: Canada-US regional trade patterns," *The American Economic Review*, 85, 615–623.

MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 71, 1695–1725.

MICHAELS, G. AND X. ZHI (2010): "Freedom fries," *American Economic Journal: Applied Economics*, 2, 256–281.

MONTALVO, J. G. AND M. REYNAL-QUEROL (2005): "Ethnic diversity and economic development," *Journal of Development economics*, 76, 293–323.

——— (2021): "Ethnic diversity and growth: Revisiting the evidence," *Review of Economics and Statistics*, 103, 521–532.

MORALES, E., G. SHEU, AND A. ZAHLER (2019): "Extended gravity," *The Review of economic studies*, 86, 2668–2712.

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

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

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

NAGAVARAPU, S. AND S. SEKHRI (2016): "Informal monitoring and enforcement mechanisms in public service delivery: Evidence from the public distribution system in India," *Journal of Development Economics*, 121, 63–78.

OECD (2020): *All hands in? Making diversity work for all*, OECD Publishing.

OH, S. (2023): "Does identity affect labor supply?" *American Economic Review*, 113, 2055–2083.

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

——— (2001): "Business and social networks in international trade," *Journal of economic literature*, 39, 1177–1203.

RAUCH, J. E. AND V. TRINDADE (2002): "Ethnic Chinese networks in international trade," *Review of Economics and Statistics*, 84, 116–130.

SANTANGELO, G. (2019): "Firms and Farms: The Local Effects of Farm Income on Firms' Demand," *Cambridge Working Papers in Economics*.

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

SPEARS, D. AND A. THORAT (2019): "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*, 67, 725–755.

STARTZ, M. (2016): "The value of face-to-face: Search and contracting problems in Nigerian trade," *Available at SSRN 3096685*.

TAKESHITA, J., S. WANG, A. W. LOREN, N. MITRA, J. SHULTS, D. B. SHIN, AND D. L. SAWINSKI (2020): "Association of racial/ethnic and gender concordance between patients and physicians with patient experience ratings," *JAMA network open*, 3, e2024583–e2024583.

TAN, E. AND T. H. ZEIDA (2023): "Consumer Demand and Credit Supply as Barriers to Growth for Black-Owned Startups," *Accepted at Journal of Monetary Economics*.

VITALI, A. (2022): "Consumer search and firm location: Theory and evidence from the garment sector in Uganda," *Work. Pap*.

A Model Derivations

Household Problem. The first-order conditions for the households are

$$\begin{aligned} \frac{a}{C_{s,H}} - \lambda P_{s,H} &= 0 \\ (1-a)\frac{q(z(\omega),s,s^{'})c(z(\omega),s,s^{'})^{\frac{-1}{\sigma}}}{C_{s,D}^{\frac{\sigma-1}{\sigma}}} - \lambda p(z(\omega),s,s^{'}) &= 0 \end{aligned} \tag{A.1}$$

where λ is the Lagrange multiplier for the constraint on the total expenditure as defined in Equation (3). This specification implies that the representative household allocates the remaining income to the two goods proportional to their weights in the utility function. Manipulating the Equation (A.1), we get that

$$P_H C_H = aI_s; \quad \sum_j p(z(\omega),s,s^{'})c(z(\omega),s,s^{'}) = (1-a)I_s; \quad \lambda = \frac{1}{I_s} \tag{A.2}$$

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.1 Firm Problem

The optimal size of the firm of quality z and firm owner from group s is given by

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

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

A.1.1 Proof Proposition 1: Optimal Location

Let $\mathcal{F}_s^d$ be the sum of all fixed costs paid by the firm. Firm-level profits of the firm owner of caste s are given by

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

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

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

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

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

Collecting terms and a few steps of Algebra give us

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

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 $\lambda_{s',s,k} = \frac{\beta\kappa_{s'} + B^{-1}_{z,s}\gamma_{z,s,k}\mathbb{1}_{s=s'}}{\Upsilon}$. $\mathbb{1}_{s=s'}$ 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 (Fermat-Weber Problem) location problem. This is a version of the Fermat-Weber Problem as our problem is quadratic in distance. This allows us to have the closed-form solution to the optimal location.

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

First-order conditions for an unconstrained firm are given by

$$-\frac{\partial}{\partial\Delta\mathcal{X}_{z,s,k}} \sum_{s'} \lambda_{s',s,k}(d_{s',s,k} - \Delta\mathcal{X}_{z,s,k})^2 = 0 \quad \forall \quad k \tag{A.9}$$

This gives us the expression for the optimal distance moved

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

For constrained firms, $\Delta\mathcal{X}_{z,s,k} = \underline{X}_{z,s,k}$

A.1.2 Derivation of Firm-level Revenues

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

$$r(z,s',s) = q(z,s',s)^{\sigma} \left(\frac{\sigma}{\sigma-1}\,\mathcal{C}_s\right)^{1-\sigma} \kappa_{s'} \tag{A.11}$$

Now, we can compute the cross-group micro trade elasticities. Taking logs on both sides and taking derivative with respect to $\log I_s$, and $\Psi^{\sigma}_{z,s',s} = e^{-\tilde{\beta}\sum_{k\in\mathcal{K}}(d^*_{ss',k})^2}$, we have

$$\frac{\partial \log r(z,s,s^{'})}{\partial \log I_s} = \underbrace{\frac{\partial \log \kappa_s}{\partial \log I_s}}_{\textbf{Size effect}} \underbrace{+2\tilde{\beta} \sum_{k\in\mathcal{K}} d^*_{ss',k} \frac{\partial\Delta\mathcal{X}_{s',k}}{\partial \log I_s}}_{\textbf{Δ Optimal Distance}} \tag{A.12}$$

A.1.3 Proof of Proposition 2

The firm-level revenues are

$$R(z,s) = \left(\frac{\sigma}{\sigma-1}\,\mathcal{C}_s\right)^{1-\sigma} z^{\sigma} \sum_{s'} \kappa_{s'} e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{K}}(d_{s's,k}-\Delta\mathcal{X}_{z,s,k})^2\right)} \tag{A.13}$$

$$\log R(z,s) = \alpha_s + \log \sum_{s'} \kappa_{s'} e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{K}}(d_{ss',k}-\Delta\mathcal{X}_{z,s,k})^2\right)}, \tag{A.14}$$

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 group s to any group s' as

$$A_{s's} = \frac{\left(\kappa_{s'} e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{K}}(d_{s's,k}-\Delta\mathcal{X}_{z,s,k})^2\right)}\right.}{\sum_{s'} \kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{K}}(d_{s's,k}-\Delta\mathcal{X}_{z,s,k})^2\right)}} \tag{A.15}$$

Taking the derivative both of the revenue equation, we get

$$\frac{\partial \log R(z,s)}{\partial \log I_{s'}} = \frac{\partial \log \kappa_{s'}}{\partial \log I_{s'}} A_{s's} + 2\tilde{\beta} \sum_{k\in\mathcal{K}} d^*_{z,s',s,k} \frac{\partial \Delta\mathcal{X}_{z,s',k}}{\partial \log I_{s'}} A_{s's}, \tag{A.16}$$

$$\frac{\partial \log R(z,s)}{\partial \log I_{s'}} = \frac{\partial \log r(z,s',s)}{\partial \log I_{s'}} A_{s's}. \tag{A.17}$$

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

$$\begin{aligned}
\frac{\partial}{\partial \log I_s} \frac{\sum_{s\in\mathcal{S}} \tilde{\lambda}_{ss',k} d_{ss',k}}{\sum_{s\in\mathcal{S}} \tilde{\lambda}_{ss',k}} =& \frac{\partial}{\partial \log I_s} \frac{\sum_{s\in\mathcal{S}} (\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'}) d_{ss',k}}{\sum_{s\in\mathcal{S}} (\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'})} \\
=& \frac{\tilde{\beta}\frac{\partial \kappa_s}{\partial \log I_s} d_{ss',k}}{\sum_{s\in\mathcal{S}} (\tilde{\beta}\kappa_s + B_{s'}^{-1}\gamma_{s',k}\mathbb{1}_{s=s'})} - \frac{\tilde{\beta}\frac{\partial \kappa_s}{\partial \log I_s} \sum_{s'\in\mathcal{S}} (\tilde{\beta}\kappa_{s'} + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'}) d_{ss',k}}{\left(\sum_{s'\in\mathcal{S}} (\tilde{\beta}\kappa_{s'} + B_{s'}^{-1}\gamma_{s,k}\mathbb{1}_{s=s'})\right)^2} \\
=& \frac{\tilde{\beta}}{Y\sum_{s\in\mathcal{S}} \tilde{\lambda}_{ss',k}} \frac{\partial \kappa_s}{\partial \log I_s} \left(d_{ss',k} - \frac{\sum_{s\in\mathcal{S}} \tilde{\lambda}_{ss',k} d_{ss',k}}{\sum_{s\in\mathcal{S}} \tilde{\lambda}_{ss',k}} \right) \\
=& \frac{\tilde{\beta}}{Y\sum_{s\in\mathcal{S}} \tilde{\lambda}_{ss',k}} \frac{\partial \kappa_s}{\partial \log I_s} d^*_{ss',k}
\end{aligned} \tag{A.18}$$

A.1.4 Proof of Proposition 3

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 group. We first formulate the general problem. The revenue shares of the firm owner of group s' are defined as

$$A_{ss'} = \frac{\kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}}{\sum_s \kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}} \tag{A.19}$$

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

$$\begin{aligned}\frac{\partial A_{ss'}}{\partial \tilde{\beta}} =& A_{ss'}\left(-\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)+2\tilde{\beta}A_{ss'}\sum_{k\in\mathcal{K}}d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}}\\ &+A_{ss'}\frac{\sum_s\kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}{\sum_s\kappa_s e^{-\tilde{\beta}\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2}}\\ &-A_{ss'}\frac{\sum_s\kappa_s e^{-\tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2\right)}2\tilde{\beta}\sum_{k\in\mathcal{K}}d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}}}{\sum_s\kappa_s e^{-\tilde{\beta}\sum_{k\in\mathcal{N}}(d^*_{ss',k})^2}}\end{aligned} \tag{A.20}$$

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

$$\begin{aligned}\frac{\partial A_{ss'}}{\partial \tilde{\beta}} =& -A_{ss'}\,(d^*_{ss'})^2+A_{ss'}\sum_s A_{ss'}(d^*_{ss'})^2+2\tilde{\beta}A_{ss'}\sum_{k\in\mathcal{K}}d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}}\\ &-2\tilde{\beta}A_{ss'}\sum_s A_{ss'}\sum_{k\in\mathcal{K}}d^*_{ss',k}\frac{\partial\Delta\mathcal{X}_{s',k}}{\partial\tilde{\beta}}\end{aligned} \tag{A.21}$$

In the case of exogenous group 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

$$\frac{\partial A_{ss'}}{\partial \tilde{\beta}} = -A_{ss'}\,(d^*_{ss'})^2+A_{ss'}\sum_s A_{ss'}(d^*_{ss'})^2 \tag{A.22}$$

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

$$\frac{\partial A_{ss'}}{\partial \tilde{\beta}} = A_{ss'}\sum_{s''}A_{s''s'}\left((d^*_{s''s'})^2-(d^*_{ss'})^2\right) \tag{A.23}$$

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 group consumer is increasing in $\tilde{\beta}$, then the firm is selling less to other groups, $1-A_{ss}$, therefore, the revenue elasticity to the income shock is lower as well.

A.2 Taste Shifters and Transportation Costs

The trade across groups may be affected by both tastes and transportation costs. 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. Here, we derive the solution to the optimal group identity under both of these distortions. We assume that physical distance (that captures transportation costs) and group distance (that captures taste) are perfectly correlated. This is a reasonable assumption since LC communities tend to be segregated.[27] Therefore firm owners from LC communities may need to include transportation costs when selling to consumers of high groups. 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 group (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

$$p(z, s', s) = \frac{\sigma}{\sigma - 1} \mathcal{C}_s \tau_{s's}, \tag{A.24}$$

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

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.[28] Further, we assume that transportation costs are quadratic in distance $\tau(d_{ss'}, \Delta d_s) = e^{\nu\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s,k})^2\right)}$, where γ disciplines the rate at which the costs increase. The overall trade barrier for a firm can be summarised 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 $\hat{\beta} = \tilde{\beta}\sigma + \nu(\sigma - 1)$ that captures the strength of the cross-group 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 groups

[27] See, for instance, Bharathi et al. (2021) for recent evidence.

[28] For example, opening a branch closer to the neighborhoods of other groups reduces transportation costs and also increases the perceived quality of the product.

declines with ethnic and physical distance, is a combination of taste parameter β and transportation cost parameter ν. We provide more details in Appendix A.[29]

We assume that 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. 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 ν.

B Empirical Analysis

B.1 Data

B.1.1 MSME dataset

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

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

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.

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, we drop if any one of the revenues, capital stock, wagebill, and the number of employees are missing. Our empirical specification exploits rainfall shocks, that affect the rural economy, thus we restrict our attention to the rural areas and non-food producing sector (NIC = 15).[30] Following Jayachandran (2006), we also restrict our attention to the major states during the time period of our data - Andhra Pradesh, Bihar, Chandigarh, Chhattisgarh, Delhi, Gujarat, Haryana, Jharkhand, Karnataka, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal. We are left with 349,715 observations. HC, MC, and LC represent 40%, 45%, and 15% of observations.

We have two definitions of contact-intensive sectors: definition 1 includes, carpentry & furniture manufacturing, construction, wholesale & retail, hotels & restaurants, activities of travel agents, post and telecommunications (this includes Provision of basic telecom services: telephone, telex and telegraph and activities of STD/ISD booths), and computer related services; definition 2 also includes repair and maintenance shops, real estate activities, renting of machinery and equipment, and education and health services.

In Table B.1, we provide the distribution for all firms including urban and rural areas; the average firm size is 6.3 employees and the majority (3.2) of the employees belong to the HC community and 1.1 employees are from the LC community. On

[30]We drop this sector, "Manufacture of food products", as it relies on the agricultural inputs that directly affected by rainfall.

average, 75% of employees belong to the same community as the firm owner – own-caste employee share. In Figure B.2, we see that there is little heterogeneity in terms of the relationship between the share of own-caste employees and firm size. Perhaps, unsurprisingly, we find that LC-owned firms face the biggest decline in the own-caste employee share. As LC constitutes the smallest share of the population and an even smaller share of income and demand, LC-owned firms have to march out to gain scale.

This relationship is further demonstrated in Table B.2, where we regress the share of own-caste employees on log(employees) and find a significant negative relationship. This is much more pronounced among MC- and LC-owned firms. Next, We find that the relationship is even stronger in contact-intensive sectors. This is in line with our hypothesis, that in sectors, where customer and employee interactions are strong, firms need to hire people from diverse backgrounds to gain market access from those groups.

All firms	Mean	Median	p5	p95	N
Emp. All	6.3	3	1	17	1360525
Emp. LC	1.1	0	0	4	1360525
Emp. MC	2.0	1	0	6	1360525
Emp. HC	3.2	0	0	1	1360525
Emp. Own-C (%)	75	100	0	100	1360525
Revenue (10^4)	470	22.5	3.5	823	1360525
Materials (10^4)	294	9.4	0.38	527	1360525

Table B.1: Firm-size distribution

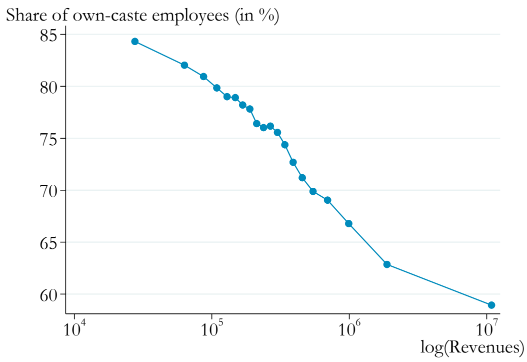
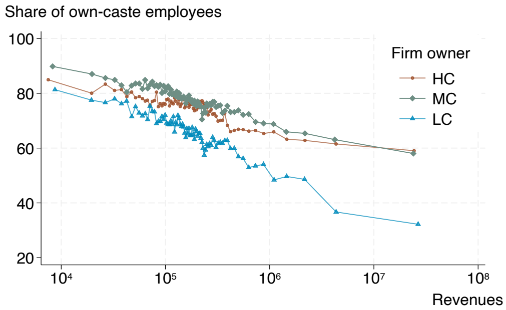


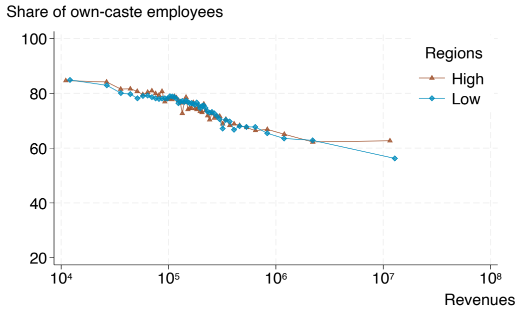
Figure B.1: Own-Caste employee share

Notes: Table B.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 B.1 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 multipliers are applied.

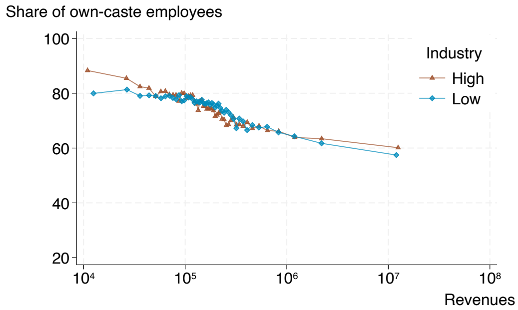
Figure B.2: Firm Size and Own-caste Employee Share



(a) By Caste



(b) By Region



(c) By Sector

Notes. The table reports binscatter plot with firm revenues on the x-axis and the share of own-caste employees on the y-axis. LC, MC, and HC represent entrepreneurs historically classified as belonging to the SC & ST, middle-castes, and high-castes, respectively. Sampling multipliers are applied in all the panels.

Table B.2: Own Caste Employee Share and Firm Size

	(1)	(2)	(3)	(4)
Outcome	Own Caste Employee Share			
log(Employees)	-6.783***	-6.453***	-5.856***	-5.414***
	(0.310)	(0.355)	(0.422)	(0.093)
log(Employees) × MC			-1.685***	-1.934***
			(0.571)	(0.132)
log(Employees) × LC			-2.938***	-2.466***
			(0.791)	(0.232)
log(Employees) × Contact-intensive sector		-1.261***		-0.107
		(0.429)		(0.970)
log(Employees) × MC × Contact-intensive sector				-1.337***
				(0.216)
log(Employees) × LC × Contact-intensive sector				-4.701***
				(0.443)
Observations	346,539	346,539	346,539	346,539
R-squared	0.182	0.182	0.183	0.185

Notes. LC, MC, and HC represent entrepreneurs historically classified as belonging to the SC & ST, middle-castes, and high-castes, respectively. State, Sector, and Caste FE are included. Sampling multipliers are applied. Robust Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

B.1.2 Caste Employment Shares by Sector

Table B.3 describes the labour share of workers in MSME firms across different sectors. The overall employment share of LC workers remains around 37 percent which is slightly above their population share. In contact-intensive sectors, we expect castes with less appeal to the firm to have a lower labour share than in sectors that are less contact-intensive, as firms try to project themselves closer to the wealthier castes. Consistent with this, Columns 1-2 show that LC labour share is higher in less contact-intensive sectors while HC labour is higher in more contact-intensive sectors. On the other hand, when it comes to sectors where LC households display high elasticities following the rainfall shock, we observe the opposite pattern. LC labour share is higher in sectors where LC consumers show a higher consumption than in other sectors, while HC labour share is substantially lower in sectors where LC consumers show a higher consumption than in other sectors.

Table B.3: Caste Employment Shares by Sector

	Contact-intensive Sector 1		Contact-intensive Sector 2		LC's high-consumption sectors	
	Yes	No	Yes	No	Yes	No
LC	0.155	0.177	0.146	0.182	0.194	0.180
MC	0.468	0.448	0.472	0.444	0.442	0.282
HC	0.378	0.375	0.382	0.373	0.364	0.539

Notes. The table reports descriptive statistics. LC, MC, and HC represent workers historically classified as belonging to the historically disadvantaged castes, middle-castes, and historically privileged castes, respectively. Sampling multipliers are applied. It reports statistics for the NSS Employment and Unemployment data set. S.D. is the standard deviation. Each row reports labour share statistics for HC, MC, LC, respectively, for contact-intensive sectors vs other sectors, and LC's high-consumption sectors vs other sectors.

B.1.3 Employment and Unemployment Data

We use data from the National Sample Survey (NSS) Employment and Unemployment survey to collect information on workers, their wages, and their demographics, at the district level. We use five schedules spanning the years 2003-04 to 2009-10. Specifically, the analysis includes the NSS schedules 60, 61, 62, 64 and 66. We use the total earnings as a measure of individual wage. This includes daily wage and contractual salary.

We then divide it by the number of days worked to obtain our variable of interest: daily wage of an individual. We use the NIC (4-digit) code of the most recent job to determine the sector of employment.[31] The descriptive statistics are provided in Table B.4. HC workers account for 28.3% of the employment survey sample, while LC workers account for 36.1% of the sample. However, Table B.4 shows that the HC workers report, on average, a higher daily wage (148.42 Indian rupees) compared to LC workers (59.47 Indian rupees), and higher education.

Table B.4: Summary statistics

Panel A: NSS 2004-2010 - Individual-level statistics								
	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Wages	144.27	(310.16)	75.87	(104.76)	56.37	(74.99)	84.63	(174.11)
Sex (male)	0.81	(0.39)	0.73	(0.44)	0.70	(0.46)	0.74	(0.44)
Age	35.18	(12.09)	34.88	(12.47)	34.58	(12.59)	34.83	(12.43)
Education	6.83	(3.94)	4.87	(3.61)	3.88	(3.26)	4.96	(3.74)
Land owned	135.05	(802.00)	133.12	(616.53)	95.22	(404.00)	118.74	(596.99)
Employed in agri.	0.26	(0.44)	0.47	(0.50)	0.59	(0.49)	0.47	(0.50)

Panel B: NSS 2004-2008 - Household-level statistics								
	HC		MC		LC		All	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
MPCE	1216.4	(1289.6)	777.5	(780.6)	602.3	(440.8)	851.5	(920.5)
Education	7.42	(4.13)	5.43	(3.93)	4.39	(3.71)	5.70	(4.10)

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 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 B 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).

[31]The survey includes questions about the activities of individuals during the most recent seven days.

B.1.4 Household Consumption Data

We use four schedules spanning the years 2003-2004 to 2007-08; 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. The descriptive statistics are provided in Table B.4 about MPCE, education of the head of the household.

B.1.4.1 Identity Engel curves: We define the social distance of a product from a consumer point of view in two steps. First, we use the consumption survey to compute the MPCE for seven categories of products: Food, Non-food, Fuel & Light, Clothing, Footwear, Durables, and Miscellaneous.[32] In the second step, we use the MSME data to define the share of employees employed in each product category by district. We define three measures of perceived social distance, each lying between 0 and 1.

The first measure is Social Distance $1_{pd} = \frac{EMP_{LC}}{EMP_{LC}+EMP_{HC}}$, which is the ratio of the share of LC employees to the sum of the shares of LC and HC employees in product markets. Our hypothesis is that as Social Distance 1_{pd} increases, the expenditure of LC consumers declines relative to HC consumers. Formally, we run the regression:

$$x_{ip} = \alpha_0 + \alpha_1 \text{Social distance}_{pd} \times caste_s + \Gamma_i \text{controls}_i + \alpha_{sp} + \alpha_{ds} + \alpha_{pd} + \epsilon_{ip}, \quad \text{(B.25)}$$

where i is a household, p is a product category, s is the caste, and d is a district. x_{ip} is log(MPCE) at the product level. We control for wealth and overall expenditure (as a proxy for income) and saturate the regression model with district $\times$ product, district $\times$ caste, and caste $\times$ product fixed effects. We expect α_1 to be negative. The results are presented in Table B.5.

We find that as the HC dominates the production and selling of a product cat-

[32] One can have more disaggregated product categories, but then one runs the risk of not having enough observations across castes and districts.

egory relative to LC, the LC consumption share decreases by 12.6% relative to HC consumers. In our second measure, we find that as MC communities dominate the product markets relative to LC, the LC consumption share decreases by 13.3% relative to MC consumers. Finally, in the third measure, we find that when HC dominates the product market relative to MC, MC consumers' consumption drops by 15.6% relative to HC consumers. Overall, this suggests that taste for identity influences consumption patterns among Indian households.

Table B.5: Identity Engel Curves

Outcome	(1) log(MPCE)	(2) log(MPCE)	(3) log(MPCE)
LC × Social Distance 1 $_{pd}$	-0.126**		
	(0.055)		
LC × Social Distance 2 $_{pd}$		-0.133**	
		(0.054)	
MC × Social Distance 3_{pd}			-0.156**
			(0.072)
Observations	84,266	98,570	108,609
R-squared	0.649	0.632	0.637
Product × District FE	✓	✓	✓
District × Caste FE	✓	✓	✓
Product × Caste FE	✓	✓	✓

Notes. We define a market pd for a product p and district d. Let EMP_{LC}, EMP_{MC}, and EMP_{HC} be the share of LC, MC, and HC employees in a market pd and Social Distance 1 $_{pd} = \frac{EMP_{HC}}{EMP_{LC}+EMP_{HC}}$, Social Distance 2 $_{pd} = \frac{EMP_{MC}}{EMP_{LC}+EMP_{MC}}$, and Social Distance 3 $_{pd} = \frac{EMP_{HC}}{EMP_{MC}+EMP_{MC}}$. We control for wealth and income for all regressions. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at the district level in all regressions, *** p<0.01, ** p<0.05, * p<0.1.

B.1.5 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 total rainfall registered on the grid points within the boundary of a district. Temporally, we aggregate this data as the total annual rainfall to construct a district-wise time series of rainfall received across Indian districts since the year 1950. We define a positive shock if the annual rainfall measure is above the 80th percentile and a negative shock if the rainfall is below the 20th percentile within the district. We drop the top 1

	log *MPCE*
Rainshock × MC	0.074***
	(0.022)
Rainshock × LC	0.088***
	(0.027)
Rainshock × Headmeals	-0.038*
	(0.022)
Rainshock × Headedu	0.013***
	(0.003)
Rainshock × land	0.002
	(0.004)
Observations	117,742
R-squared	0.328
Controls$_{it}$	✓
Caste FE	✓
District × *Year* FE	✓

Table B.6: Consumption

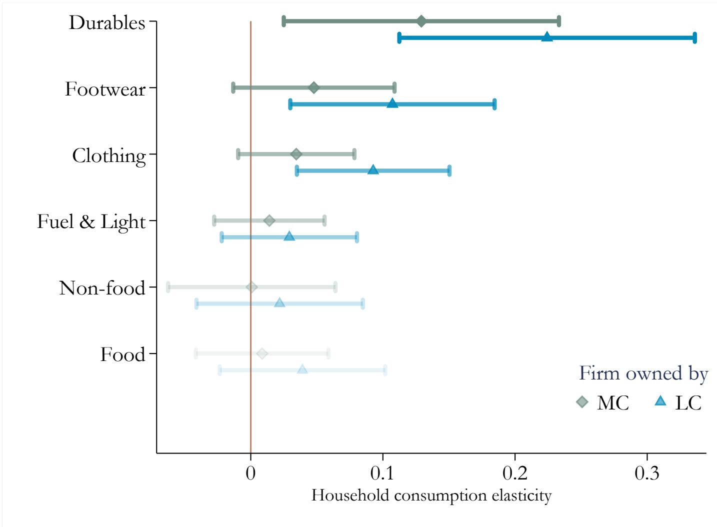


Figure B.3: Heterogeneity in consumption

Notes. Table 2 presents the elasticity of monthly per-capita consumption of households in our sample. The regressions are of household-level variables (logarithmic) on rainfall shock by caste. Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at the district level in all regressions, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Figure 4 presents the heterogeneity in consumption elasticities with values on the x-axis and the consumption groups labeled on the y-axis. We control for household-level wealth (using meals per day and land owned) and education, and their interactions with rainfall. We also control for caste, district, and year fixed effects.

percentile of districts with excessive rainfall to avoid cases of floods. For the analysis, we follow Jayachandran (2006) and define rain shock as equal to +1 for positive shock, -1 for negative shock, and 0 otherwise. Following Adhvaryu et al. (2013), we focus on the 16 largest Indian states which account for nearly 95 percent of India's population.

B.2 Robustness: Consumption Elasticity

Table B.6 and Figure B.3 present consumption elasticities, controlling for the household head's meals and land ownership (as proxies for wealth), as well as education interacted with rainfall shocks. The results are consistent with the baseline estimates, indicating that the wealth-based explanation, where only poor households respond to rainfall shocks, cannot fully account for the findings. Land ownership has an insignificant effect, while education has a positive effect. Additionally, we find a negative association between rainfall shocks and meals per day, suggesting that wealthier households respond less than poorer ones. However, the persistent caste effect implies that the wealth channel is not the key driver of rainfall-induced demand shocks.

B.3 Geographical segregation

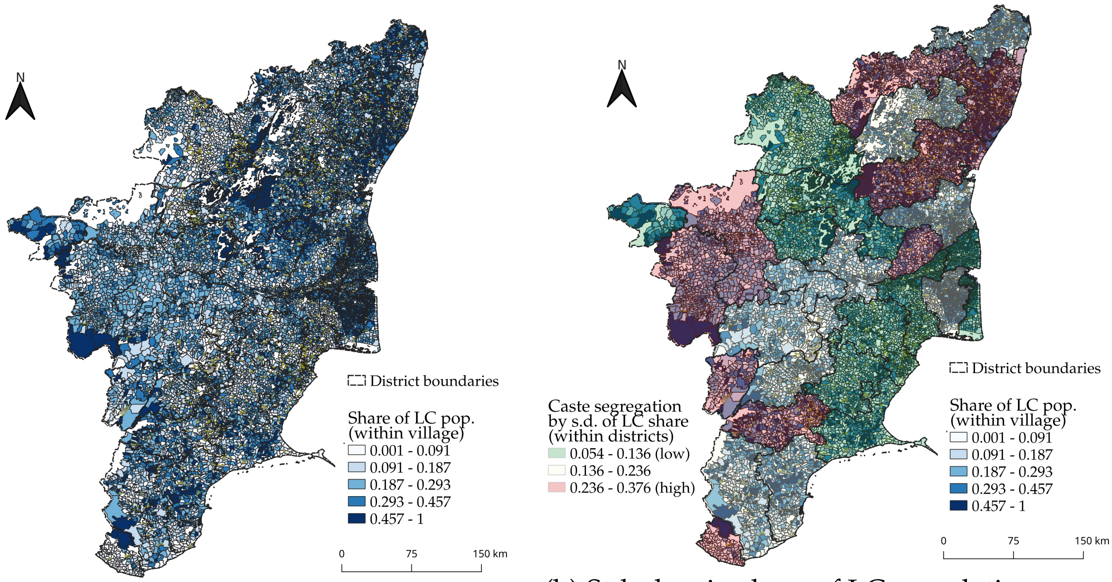
The premise of our study is to investigate whether firms belonging to a certain ethnic group tend to cater to consumers of the same group, and hence suffer from the consequences to growth due to limited demand. In Section 5, we have documented the presence of such ethnic linkages and the resulting segregation 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 are geographically farther.[33]

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 SHRUG database (Asher and Novosad, 2019). Within each village, we obtain the share of LC (=SC+ST) population. To measure geographical segregation across castes within the district, we use the standard deviation in the share of LC population in villages within a district; $\text{Segregation}_d = sd(LCshare_v)$, where d denotes district and v denotes village and sd denotes the standard deviation function.

Figure B.5 shows that there is a lot of variation in within-district geographical segregation across India. Figure B.4 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 or decile of our measure of segregation. Table B.7 shows that the results are in line with our baseline results from Table 3.

[33] Recent papers (see, e.g., Jensen and Miller, 2018) have shown that geographical distance can be an important factor in determining the market of a firm's product.

Figure B.4: 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.

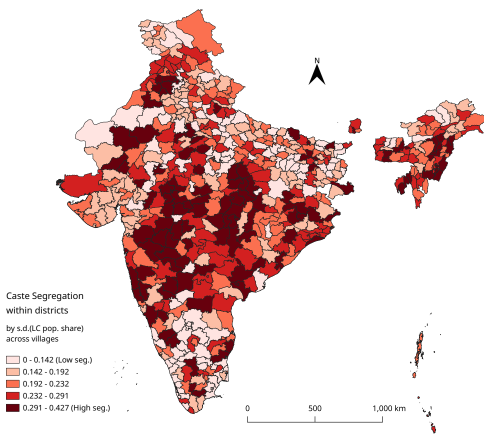


Figure B.5: Geographical segregation, by caste, across districts of India (2001)

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.7: Elasticities in *unsegregated* rural India, in LC's High-consumption Sectors

Segregated districts:	top quartile				top decile			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	Revenues	Inputs	Revenues	Inputs	Revenues	Inputs	Revenues	Inputs
Rainshock × MC	0.048	0.045	0.050	0.046	0.079*	0.100	0.082**	0.099*
	(0.051)	(0.068)	(0.044)	(0.060)	(0.046)	(0.062)	(0.040)	(0.054)
Rainshock × LC	0.152*	0.209**	0.126**	0.180**	0.193***	0.280***	0.165***	0.242***
	(0.080)	(0.100)	(0.057)	(0.078)	(0.070)	(0.090)	(0.053)	(0.072)
Observations	267,756	264,889	266,607	263,727	322,599	319,384	321,502	318,274
R-squared	0.512	0.519	0.594	0.587	0.507	0.515	0.585	0.578
Caste FE	✓	✓	✓	✓	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Sector × Year	✓	✓			✓	✓		
Product × Year FE			✓	✓			✓	✓

Notes. The regressions are of firm-level variables (logarithmic) on rainfall shock by caste. Columns 1 to 4 include observations from all sectors. Columns 5 to 8 include observations from the sectors where LC households display high elasticities following the rainfall shock. We control for caste, district × year, and sector × year or product × year fixed effects. Sample omits observations of districts with standard deviation in village-level LC population share in the top quartile (Columns 1 to 4) and top decile (Columns 5 to 8). 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.

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.

B.4 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 demand constraints of an LC firm. Could the results on firm growth be explained by rise in prices due to increase in demand? Could the results on firm growth be explained by reasons such as relaxation of credit 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.

First, we estimate the effect of higher rainfall on the following firm-level variables

Table B.8: Cross-sectional firm-level elasticities in rural India: MSME 2006-07

VARIABLES	(1) Output	(2) Material Input	(3) Output Price	(4) Input Price	(5) All	(6) Institutional	(7) Non-institutional	(8) log($mrpk$)
MC	-0.304***	-0.374***	-0.020	-0.162***	-0.481***	-0.883***	-0.834***	0.026
	(0.025)	(0.030)	(0.016)	(0.052)	(0.132)	(0.162)	(0.318)	(0.023)
LC	-0.523***	-0.653***	-0.042*	-0.302***	-0.574***	-0.878***	-1.061***	0.081***
	(0.029)	(0.038)	(0.025)	(0.068)	(0.092)	(0.158)	(0.366)	(0.028)
Rainshock × MC	0.080*	0.122*	-0.048*	-0.044	0.124	0.156	0.500	-0.029
	(0.043)	(0.067)	(0.026)	(0.123)	(0.189)	(0.237)	(0.629)	(0.039)
Rainshock × LC	0.139***	0.234***	-0.013	-0.005	0.180	0.242	0.365	-0.073
	(0.050)	(0.072)	(0.045)	(0.117)	(0.171)	(0.344)	(0.591)	(0.049)
Observations	335,666	333,612	482,430	107,212	123,792	26,707	1,825	335,434
R-squared	0.596	0.610	0.766	0.419	0.506	0.456	0.641	0.433
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Columns (1) to (2) show the regressions of firm-level variables (logarithmic) on rainfall shock in the year 2006-07. Columns (3) and (4) show the regressions of product-level prices on rainfall shock in the year 2006-07. Columns (5) to (8) show the regressions of firm-level loans taken (logarithm of 1+loan value) and $mrpk$ (logarithmic) on rainfall shock in the year 2006-07. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

(y): (1) revenue, and (2) material input. 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.

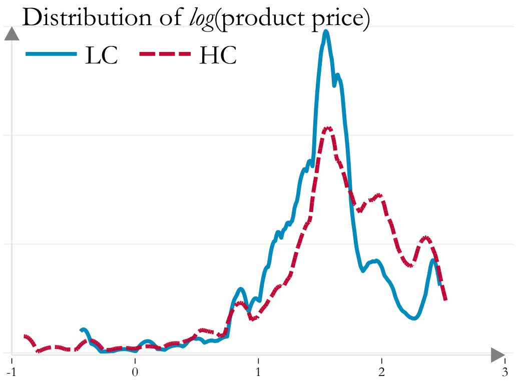
Third, we estimate the effect on the following firm-level variables (y): (1) all loans taken, (2) institutional loans taken, (3) non-institutional loans taken, and (4) marginal revenue product of capital (MRPK). This exercise sheds light on two issues: to check whether the change in firms' outcomes in the previous regressions 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.6.

In Table B.8, 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.

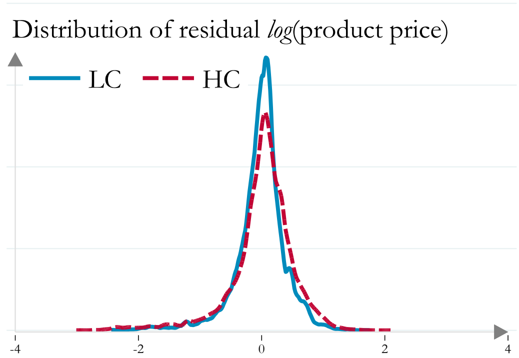
Prices. To further explore the mechanism behind the effect of higher rainfall on firm revenue, we estimate the effect of higher rainfall on input and output prices. Table B.8, Columns (3) and (4) show that there is no significant difference in the average prices of LC-owned firms relative to HC-owned firms. Figure B.6 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-owned 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, is not through its effect on input or output prices.

Credit and Financial Frictions. 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.8 shows that value of loans LC-owned firms have taken is lower than that of HC-owned firms, in a district with median rainfall.[34] 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 shocks affect firms, that is through formal (e.g., banks) or informal (e.g., caste networks) credit sources. An alternative outcome that may capture the relaxation of credit constraints 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.8, Column (9) shows no significant change along these lines. These strings of evidence suggest that loosening of credit constraints is not driving the firm growth.

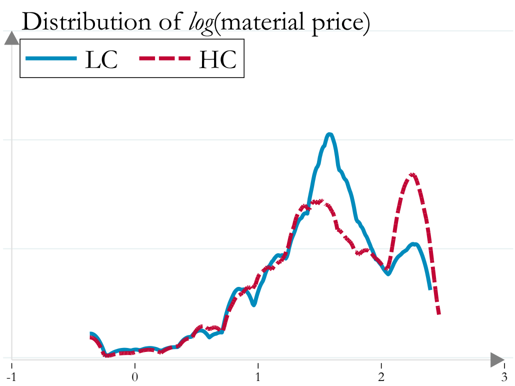
[34] Goraya (2023) shows that these differences are not driven by productivity, and establishes the existence of misallocation across caste due to credit-constraints.



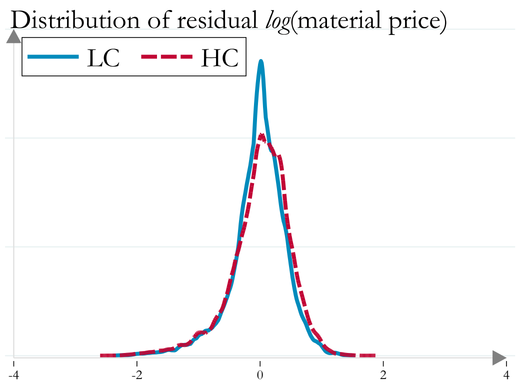
(a) output prices: unconditional



(b) output prices: within District and Product-Category



(c) input prices: unconditional



(d) input prices: within District and Product-Category

Figure B.6: Distribution of output and input 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.

Firm characteristics. In Table B.9, 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.9, 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.

Table B.9: Heterogeneity in firm-level elasticities in rural India: MSME 2004-07

Mediating variable (*hetvar*):	Small firms		Top LC sectors	
	(1)	(2)	(3)	(4)
Outcome:	Revenue	Material Input	Revenue	Material Input
Rainshock × MC	0.211***	0.290***	0.115***	0.158**
	(0.054)	(0.082)	(0.041)	(0.062)
Rainshock × LC	0.244***	0.357***	0.140***	0.210***
	(0.070)	(0.094)	(0.047)	(0.066)
hetvar	-1.593***	-2.097***	-1.043	-0.351
	(0.362)	(0.426)	(1.040)	(1.140)
Rainshock × *hetvar*	-1.653***	-2.001***		
	(0.050)	(0.065)		
MC × *hetvar*	0.374***	0.501***	-0.040	-0.066
	(0.048)	(0.068)	(0.060)	(0.078)
LC × *hetvar*	0.443***	0.548***	-0.177	-0.072
	(0.065)	(0.090)	(0.108)	(0.131)
Rainshock × MC × *hetvar*	-0.159**	-0.222**	-0.012	0.007
	(0.062)	(0.092)	(0.086)	(0.116)
Rainshock × LC × *hetvar*	-0.164**	-0.270***	-0.116	-0.158
	(0.071)	(0.093)	(0.089)	(0.119)
Observations	950,345	941,873	950,345	941,873
R-squared	0.624	0.626	0.512	0.544
District × Year FE	✓	✓	✓	✓
Product × Year FE	✓	✓	✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall shock. Small firms indicate firms with below-median revenue, within caste-categories. Top LC sectors are the top five 4-digit sectors 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.

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.9, Columns (3) and (4) show a lack of any significant difference in the revenue and material input, between the top five 4-digit sectors for LC firms (by revenue) and other sectors. It also shows that the effect of rainfall is not driven by these industries.

Product characteristics. Apart from the geography-based and firm characterics-based approach, we further take a product characteristics-based approach to investigate the concern of whether the goods consumed by different castes are segregated, and whether the LC firms that show growth are indeed competing with MC and HC firms. To do this, in Table B.10, we restrict the sample to homogeneous goods. To define homogeneous goods, we use the definition from Rauch (1999) (i) conservative measure in Columns (1) and (2), and (ii) liberal measure in Columns (3) and (4). The table shows no significant difference in the prospects of the firms operating in the homogeneous-goods sectors. It also shows that the effect of rainfall on firms is solely driven by information frictions.

Table B.10: Firm-level elasticities among homogeneous goods

	(1) Revenues	(2) Inputs	(3) Revenues	(4) Inputs
Rainshock × MC	0.160***	0.208***	0.142***	0.185***
	(0.053)	(0.073)	(0.041)	(0.058)
Rainshock × LC	0.232**	0.322***	0.164***	0.230***
	(0.095)	(0.122)	(0.048)	(0.068)
Observations	157,936	157,117	156,883	156,064
R-squared	0.402	0.414	0.525	0.525
Caste FE	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓
Sector × Year	✓	✓		
Product × Year FE			✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall shock. Homogeneous goods are as defined by Rauch (1999)'s conservative measure. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

Firm Outcome among products with low price dispersion In Table B.11, we restrict our attention to products where dispersion in prices below median. This exercises addresses concerns on whether the demand is driven by products where information frictions are dominant. We find that firms' outcomes are qualitatively similar to the baseline Table B.8. This evidence adds support to our claim that consumption is segmented across castes even in fairly homogeneous goods .

Table B.11: Firm-level elasticities in rural India, for industries with low price dispersion

Outcome:	(1) Revenues	(2) Inputs	(3) Revenues	(4) Inputs
Rainshock × MC	0.072*	0.066	0.059*	0.048
	(0.038)	(0.054)	(0.030)	(0.044)
Rainshock × LC	0.090**	0.110*	0.081**	0.096**
	(0.043)	(0.057)	(0.032)	(0.045)
Observations	597,780	591,564	597,293	591,069
R-squared	0.529	0.561	0.599	0.616
Caste FE	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓
Sector × Year	✓	✓		
Product × Year FE			✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall shock. The sample is restricted to only those industries where price dispersion is below median. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

Firm outcome among products with similar quality across castes In Table B.12, we focus on products where LC- and HC-owned firms produce similar quality goods, using input price data as a proxy for product quality. Since input usage is only available at the firm level (not for individual products), we assume the firm's input mix reflects the quality of the primary product it sells. This is a reasonable approximation, as these firms are relatively small and typically do not produce or sell many different products. We calculate the average price paid by each firm for inputs, using input expenditure shares to compute a firm-level input price. Next, we calculate the average input price at the product-by-caste level and examine the difference between LC- and HC-owned firms. We then restrict our analysis to product markets where this price difference is below the median. This approach addresses concerns that the revenue increase for LC firms may be driven by their specialization in low-quality goods. Our findings show that the revenue increase for LC-owned firms is quantitatively similar to, or even greater than, the baseline results in Table 3. This supports the claim that consumption is segmented across castes, independent of product quality.

Table B.12: Firm-level elasticities in rural India, for Similar Quality Products

Outcome:	(1) Revenues	(2) Inputs	(3) Revenues	(4) Inputs
Rainshock × MC	0.076**	0.090*	0.084**	0.100**
	(0.038)	(0.051)	(0.032)	(0.045)
Rainshock × LC	0.121**	0.166**	0.121***	0.162***
	(0.050)	(0.066)	(0.042)	(0.055)
Observations	431,055	426,868	431,021	426,834
R-squared	0.504	0.510	0.557	0.556
Caste FE	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓
Sector × Year	✓	✓		
Product × Year FE			✓	✓

Notes. The regressions of firm-level variables (logarithmic) on rainfall shock. The sample is restricted to only those industries where price dispersion is below median. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

B.5 Foreign Demand Shocks

The MSME survey provides data on the value of exports in the years 2006 and 2007. This allows us to create a balanced panel of firms. We focus on rural exporting firms and their respective products to maintain compatibility with our caste-specific demand shocks. To compute foreign demand shocks for these products, we use international trade flows from CEPII's BACI dataset, which reports values of bilateral trade flows at the 6-digit Harmonized System (HS) product classification level. For each HS product code, we compute a foreign demand shock faced by exporting firms (explained below), and then merge these shocks to MSME product codes. We created a cross-walk between MSME product codes and HS product codes by hand. For precision, we focus only on the manufacturing sectors, as product descriptions are relatively similar. In the end, we are able to match 1092 out of 1688 products (65%) exported by firms in our sample. We end up with approximately 8,500 exporting firms and 12,000 observations for which we have a foreign demand shock.[35]

Consider an exporter that produces a product p at time t. We observe over the entire period products p are sold in destinations $d \in \Xi_d$, where Ξ_d is the set of all destinations India exports p to. Let X_{dpt}^{-I} denote the aggregate import flow of product

[35]Our sample captures 66% of the rural exporting firms in our sample.

p into destination d from all countries except India at time t. Thus, X_{dpt} reflects the size of the foreign market for product p at destination d at time t. The intuition we pursue below is that subsequent changes in destination d's imports of product p from the world (except India) serve as a good proxy for the change in export demand faced by Indian firms operating in market p. By leaving India's exports out of X_{dpt}, we diminish the impact of supply side effects that may also affect Indian exports. We then compute the year-to-year change in (p, d) demand as the growth rate and sum across destinations d weighted by the base-year relative importance of destination d for Indian firms:

$$FD_{pt} = \sum_{d \in \Xi_d} \overline{s_{dp}} \cdot \Delta \log X_{dpt}^{-I}. \tag{B.26}$$

where FD denotes our measure of foreign demand-shock for product p at time t on Indian firms, $\Delta \log X_{dpt}^{-I}$ is the growth in the aggregate import demand for product p at destination d, and $\overline{s_{dp}} \equiv \frac{X_{dp}^{I}}{\sum_g X_{gp}^{I}}$ denote weights constructed using baseline values of imports of product p to destination d from India. The weights are the ratio of the destination-specific trade to the total trade of product p, averaging over the sample period to alleviate the endogeneity problem. Following Barrows and Ollivier (2021), we compute growth rates as $\left(\frac{X_{dpt}^{-I} - X_{dpt-1}^{-I}}{0.5\left(X_{dpt}^{-I} + X_{dpt-1}^{-I}\right)}\right)$, to deal with situation when trade flows switched from zero to a positive number (a common feature of international trade data).

Table B.13 presents the results. Columns 1 to 3 use revenues, inputs and exports as the outcome variables, from our panel data. We find that a 1 percentage point increase in foreign demand growth led to an increase in exports by 2.2 percent and revenues by 0.58 percent. The cross-sectional results are also qualitatively similar – revenues and employment respond positively to a foreign demand shock. However, the coefficients of the LC employee share are quantitatively small and insignificant. These results suggest that foreign demand shocks, unlike rainfall shocks, are caste neutral and therefore do not lead to any changes in LC employee share across HC- or LC-owned firms. Further, it also goes against the hypothesis that firms always hire LC employees in response to temporary demand shocks.

Table B.13: Foreign Demand Shocks and Firm Outcomes

	Panel 2006-2007			Cross-section 2007				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	Exports	Revenues	Inputs	Revenues	Employment	LC Share	LC Share	LC Share
ΔFD_{pt}	2.223***	0.582*	0.808**	1.934***	1.023**	0.050	0.046	-0.004
	(0.794)	(0.343)	(0.374)	(0.703)	(0.416)	(0.075)	(0.103)	(0.286)
Observations	12,163	12,163	11,906	3,930	3,930	3,930	2,862	357
R-squared	0.705	0.765	0.744	0.879	0.848	0.644	0.610	0.542
District × year FE	✓	✓	✓					
Sector × year FE	✓	✓	✓					
Caste × year FE	✓	✓	✓					
Product FE	✓	✓	✓					
District × Sector × Caste FE				✓	✓	✓	✓	✓

Notes. Columns (1) to (3) present results for the panel data, where we have information on revenues, exports and inputs. Columns (4) to (8) present results for the cross-section of 2007, where we have information on employment and employee caste shares. The dependent variable in Column (1) is log(value of export), in Column (2) and (4) is log(revenues), in Column (3) is log(input) and in Column (5) is log(employment). The dependent variable in Columns (6), (7) and (8) is LC employee share: Column (6) considers all firms, Column (7) only considers the sample of HC-owned firms, and Column (8) only considers the sample of LC-owned firms. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

B.6 Robustness Checks

B.6.1 Robustness of wage patterns to worker sample

Table B.14 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.

B.6.2 Robustness of firm-level patterns to aggregate wealth status of households

Similarly, one may be concerned that consumption patterns of poor households drive the change in firms' outcomes, and not a result of caste. To alleviate this concern, we repeat the exercise from Table B.15, by controlling for land-owned, education, meals consumed by the head of household at the district×caste×year level, interacted with rainfall. We find that firms' outcomes in Table B.15, Columns (1) and (2) are qualita-

Table B.14: Wage elasticity in rural India for full sample of workers: NSS 2004-10

	(1) All	(2) Agri.
Rainshock	0.011	-0.027
	(0.021)	(0.024)
MC	-0.201***	-0.084***
	(0.015)	(0.021)
LC	-0.240***	-0.093***
	(0.016)	(0.018)
Rainshock × MC	0.023	0.057**
	(0.021)	(0.026)
Rainshock × LC	0.013	0.057**
	(0.026)	(0.025)
Observations	192,425	73,003
R-squared	0.493	0.322
District FE	✓	✓
Year FE	✓	✓
Controls	✓	✓

Notes. The regressions of individual level wage (logarithmic) on rainfall shock. *Agri.* stands for agricultural workers 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, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

tively similar to the baseline Table 3. Further, Table B.15, Columns (3) to (8) show that the results in Tables B.8 are also robust to this concern.

B.6.3 Robustness to alternative clustering of standard errors

In Table B.16, 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.6.4 Regional Caste Composition and Firms size

We use the population census of 2001 to compute the Caste Herfindahl-Hirschman Index, Caste HHI, of each district. We keep all states to preserve the maximum amount of observations. The firm size is positively correlated with caste concentration – less diversity – and this correlation is even higher for the contact-intensive sector, see Table B.17. In panel A, we use population census and in Panel B, we use NSS household survey to compute the caste HHI. The population census only gives information on SC, ST and the rest, while NSS provides information on OBC as well.

Table B.15: Firm-level elasticities in rural India, controlling for aggregate wealth status

VARIABLES	(1) Output	(2) Material Input	(3) Output Price	(4) Input Price	(5) All	(6) Institutional	(7) Non-institutional	(8) log($mrpk$)
MC	-0.325***	-0.398***	-0.030*	-0.177***	-0.436***	-0.405***	-0.111	0.012
	(0.025)	(0.031)	(0.017)	(0.052)	(0.070)	(0.049)	(0.154)	(0.024)
LC	-0.533***	-0.664***	-0.050*	-0.326***	-0.590***	-0.575***	-0.501**	0.069**
	(0.030)	(0.040)	(0.027)	(0.070)	(0.112)	(0.087)	(0.241)	(0.029)
Rainshock × MC	0.104**	0.147**	-0.032	-0.069	-0.236*	-0.169*	-0.264	-0.020
	(0.046)	(0.069)	(0.031)	(0.112)	(0.130)	(0.099)	(0.293)	(0.042)
Rainshock × LC	0.153***	0.258***	0.003	-0.048	0.126	-0.051	-0.456	-0.084
	(0.054)	(0.081)	(0.048)	(0.113)	(0.137)	(0.106)	(0.344)	(0.054)
Observations	335,332	333,279	481,881	107,090	23,534	19,456	1,036	335,100
R-squared	0.596	0.610	0.766	0.420	0.477	0.519	0.712	0.433
District FE	✓	✓	✓	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes. Columns (1) to (2) show the regressions of firm-level variables (logarithmic) on rainfall shock in the year 2006-07. Columns (3) and (4) show the regressions of product-level prices on rainfall shock in the year 2006-07. Columns (5) to (8) show the regressions of firm-level loans taken and $mrpk$ (logarithmic) on rainfall shock 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.

Table B.16: Elasticities in rural India, alternative clustering of standard errors

	All Sectors				LC's High-consumption Sectors			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Outcome:	Revenues	Inputs	Revenues	Inputs	Revenues	Inputs	Revenues	Inputs
Rainshock × MC	0.116**	0.162**	0.103**	0.138**	0.094	0.121	0.092*	0.115*
	(0.048)	(0.066)	(0.040)	(0.055)	(0.057)	(0.080)	(0.049)	(0.068)
Rainshock × LC	0.134**	0.200**	0.121**	0.176**	0.184**	0.268**	0.155**	0.227**
	(0.063)	(0.088)	(0.051)	(0.071)	(0.083)	(0.112)	(0.065)	(0.093)
Observations	950,345	941,873	947,614	939,134	407,531	403,971	406,517	402,952
R-squared	0.512	0.544	0.594	0.610	0.463	0.471	0.543	0.539
Caste FE	✓	✓	✓	✓	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Sector × Year	✓	✓			✓	✓		
Product × Year FE			✓	✓			✓	✓

Notes. The regressions are of firm-level variables (logarithmic) on rainfall shock by caste. Columns 1 to 4 include observations from all sectors. Columns 5 to 8 include observations from the sectors where LC households display high elasticities following the rainfall shock. We control for caste, district × year, and sector × year or product × year fixed effects. 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 state-year level in all regressions, *** p<0.01, ** p<0.05, * p<0.1.

Table B.17: Firm Size and Regional Caste Concentration

Outcome	(1) log(Emp)	(2) log(Revenues)	(3) log(Capital)	(4) log(Emp)	(5) log(Revenues)	(6) log(Capital)
Panel A. HHI computed with Population Census						
Caste HHI	0.409***	0.858***	0.560**	0.216*	0.466*	0.108
	(0.120)	(0.248)	(0.259)	(0.117)	(0.248)	(0.259)
Contact-Intensive Sector × Caste HHI				0.378***	0.768***	0.888***
				(0.135)	(0.286)	(0.298)
Observations	1,170	1,170	1,170	1,170	1,170	1,170
R-squared	0.422	0.477	0.573	0.628	0.650	0.716
State FE	✓	✓	✓	✓	✓	✓
Panel B. HHI computed with Household Survey Data						
Caste HHI	1.568**	2.483*	1.671	1.197**	2.009	0.482
	(0.658)	(1.357)	(1.430)	(0.596)	(1.247)	(1.321)
Contact-Intensive Sector × Caste HHI				0.477	0.417	1.824
				(0.552)	(1.155)	(1.224)
Observations	1,163	1,163	1,163	1,163	1,163	1,163
R-squared	0.407	0.458	0.564	0.618	0.641	0.708
State FE	✓	✓	✓	✓	✓	✓

Notes. The regressions are of district-level variables. Caste HHI is computed using the caste population shares from the population census of 2001 in Panel A, and the household survey of 2006-07 in Panel B. We control for State fixed effects. Sampling multipliers are applied in all regressions. Robust Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.