Identity, Market Access, and Demand-led Diversification*

Sampreet Singh Goraya† Akhil Ilango‡

Abstract

How costly is identity-based market segmentation? Using Indian microdata on employer–employee caste composition and household consumption, we document that firms hire workers from other caste groups to access broader consumer markets. Exploiting rainfall-induced shifts in caste-specific demand, we show that high-caste firms in competitive markets respond to rising low-caste spending by hiring more low-caste employees. We build a quantitative trade model that separates demand-side taste bias from supply-side hiring barriers. The calibrated barriers are comparable in magnitude to geographic frictions separating Indian states. Counterfactuals show that supply-side hiring frictions are the dominant source of aggregate income losses, that the two channels are complementary, and that the burden of segmentation falls disproportionately on the poorest group, whose limited home market makes cross-group exchange most valuable.

Keywords: Identity, Market access, Firm size distribution, Hiring, Misallocation, Trade, Caste system

JEL classification: O11, L11, L25, M14

*This Draft: April 5, 2026. We thank the discussants, Aruni Mitra, Sophie Nottmeyer, and Manaswini Rao, for their insightful comments. We thank Johannes Boehm, Jonas Hjort, Amit Khandelwal, David Krisztian Nagy, Suzanna Khalifa, Kaivan Munshi, Martin Rotemberg, Karthik Sastry, and Eric Verhoogen as the seminar participants at the NYU, PSE, Mannheim, Toronto, Princeton, John Hopkins University, UCSD, ULB, Turku, UC Merced, Ashoka 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

Consider a seller from a disadvantaged caste in rural India. Higher-ranked caste consumers may bypass her shop in favour of a caste-proximate competitor offering identical goods. To attract those customers, she could hire a higher-caste employee whose presence signals social proximity—but doing so is costly, requiring search, screening, and management of a mixed-caste workforce that may erode her margins. This scenario illustrates two distinct channels through which social identity segments markets: a demand-side preference for buying from socially closer sellers, and a supply-side cost of hiring across group lines. Individually, each barrier reduces a firm's market reach and distorts its hiring decisions. Together, they interact: taste bias limits the customer base, which reduces the incentive to hire across group lines, while hiring barriers raise costs for firms attempting to reach new consumers. Collectively, these micro distortions alter the equilibrium distribution of firms, the allocation of labour, and aggregate real income. We ask: what are the costs of identity-based market segmentation for aggregate real income and cross-group inequality, and which channel—demand-side preferences or supply-side hiring barriers—matters more?

We document four empirical facts linking caste identity to hiring and demand composition, develop a quantitative trade model that separates demand-side taste bias from supply-side hiring barriers, and calibrate it using microdata from rural India. The model features heterogeneous firms that use destination-group labour—a firm selling to higher-caste consumers must hire higher-caste workers—linking demand and supply barriers within the firm's cost structure. Counterfactual analysis shows that supply-side hiring barriers are the dominant source of misallocation, and that market access is most valuable for the smallest, most disadvantaged group.

Rural India provides a natural setting to study these barriers. The majority of the population resides in rural areas, where firms are small, markets are local, and producers and consumers interact face to face—making the social identity of the seller readily observable. Caste norms fragment the population into groups with restricted cross-caste interactions. Caste is inherited at birth and largely determines an individual's social identity. Despite rapid economic development, identity remains salient in market transactions: 46.7% of villages report that Dalits (Scheduled Castes) are not allowed to sell to milk cooperatives, and 35.4% report Dalits are prevented from selling in local markets (IIDS, 2010). For the remainder of the paper, we group castes into three categories: low-ranked castes (LC), comprising Scheduled Castes and Scheduled Tribes; middle-ranked castes (MC), comprising Other Backward Classes; and high-ranked castes (HC). LC households are the poorest and have the smallest share of consumer spending—a feature that will be central to the model's predictions.

We use the 2006–07 Micro, Small, and Medium Enterprises (MSME) Survey, which uniquely records caste information for both employers and employees alongside product revenues, quantities, and detailed financial variables (Goraya, 2023). We merge the firm data with household consumption data from the National Sample Survey to measure caste-specific local demand, and exploit a balanced panel of three years for revenues and material costs. A key feature of the data is that 99% of firms operate in sectors where all three caste groups compete—no single group owns more than 80% of firms. Even in markets that were historically caste-specialised—such as custom tailoring, traditionally associated with the Darji community (OBC)—there is little evidence of extreme specialisation today: MC entrepreneurs own 53% of tailoring firms, HC own 27%, and LC own 19% (Online Appendix Table A.3). Our empirical focus is on such sectors, where all castes operate and identity-based preferences create demand segmentation.

Using the linked data, we document four facts on cross-caste hiring and demand segmentation. *First*, cross-caste hiring is prevalent but asymmetric. On the extensive margin, 38% of firms employ at least one worker from another caste, but participation is uneven: only 14% of MC-owned firms hire any LC workers, compared to 24% of HC-owned firms. On the intensive margin, own-caste workers account for 50–64% of employment—exceeding a random-hiring benchmark by 13–29 percentage points, with LC- and MC-owned firms showing the largest deviations.

Second, as firms grow they diversify their workforce: the own-caste employment share declines from about 85% among the smallest firms to roughly 58% among the largest. This diversification is not uniform—cross-caste hiring expands across all remaining caste groups, albeit at different speeds, and the size-diversity gradient is steeper among LC-owned firms, consistent with a limited home market inducing expanding firms to enter out-group markets. The gradient is also steeper in contact-intensive (face-to-face) sectors where customer taste bias is more salient.

Third, the caste composition of a firm's workforce tracks the caste composition of local consumer demand, and this relationship is significantly stronger in the food sector—where caste purity norms are most salient—relative to non-food firms in the same district. Since both types of firms draw from the same local labour pool, the amplification cannot reflect population composition and instead isolates the demand-side identity channel.

Fourth, the cross-sectional correlations in Fact 3 may reflect unobservable district characteristics. To establish causality, we exploit rainfall variation—a well-established source of exogenous demand shocks in India. Positive rainfall shocks raise LC household consumption by 9.2% relative to HC households, generating a compositional shift in local demand driven by LC households' greater exposure to agriculture and lower baseline consumption. On the firm side, LC-owned firm revenues rise by 4.7% relative to HC-owned firms in the preferred specification. Crucially, HC-owned firms in competitive and contact-intensive markets respond to rising LC demand by hiring more LC employees—the direct evidence that firms diversify their workforce to capture cross-caste consumer demand.

We show that these patterns are robust to controlling for spatial segregation, product quality (input and output prices), and financial constraints (small vs large firms). A placebo using foreign demand shocks—which are expansionary but caste-neutral—confirms that the asymmetric hiring response is specific to local, caste-differentiated demand: export-driven expansions do not produce differential hiring of LC workers.

To understand the aggregate implications of these barriers, we develop a quantitative trade model in which consumers discount varieties sold by socially distant groups and firms can reduce this distance by hiring workers from the target consumer group. The model features heterogeneous firms, monopolistic competition, and endogenous entry, with three caste groups functioning as segmented markets. Cross-group barriers enter through three channels: a demand-side taste bias that discounts cross-group varieties, a hiring-cost wedge that raises the marginal cost of employing cross-group workers, and a fixed cost of entering cross-caste markets. The hiring-cost channel is motivated by a large body of empirical evidence showing that ethnic diversity within firms reduces productivity (Hjort, 2014).[1]

The production uses labour from the destination market—a firm selling to higher-caste consumers must hire higher-caste workers—so the demand and supply barriers are tightly linked in the cost structure. This destination-labour assumption reflects the micro-foundation that serving a consumer group requires workers who understand that group's preferences and product norms, and whose social proximity to the consumer facilitates face-to-face transactions (Section 5). Each barrier maps to a specific set of empirical moments: the taste bias is identified from the revenue elasticity in Fact 4—if demand were caste-blind, a shift in LC consumer spending would raise all firms' revenues equally; the differential response of LC- versus HC-owned firms reveals the degree of demand segmentation—while the hiring wedges and fixed costs of entering cross-caste markets are matched to the worker-composition shares and hiring participation rates in Fact 1.

The calibrated taste-bias parameter is $\beta = 0.284$, implying that consumers discount cross-caste varieties by a factor $\Psi = e^{-\beta} \approx 0.753$ relative to own-caste varieties; we assume this bias is symmetric across caste pairs. Since the hiring wedges g_k and export fixed costs c_x are bilateral and asymmetric, the model still generates differential market access across groups even with a common β. The hiring-cost wedges g_k average 13.1% relative to the prevailing wage across the six bilateral pairs but vary substantially. The largest wedge is MC$\rightarrow$LC at 0.177; the smallest is LC$\rightarrow$HC at 0.047—despite

[1] Caste and religious identity create co-worker frictions (Afridi et al., 2024; Ghosh, 2024), and cross-group hiring is costlier when managers are biased (Glover et al., 2017).

the largest social distance in the caste hierarchy. This low wedge reflects the data: HC workers account for 30.06% of employment in LC-owned firms (Table 2), a surprisingly high share that runs contrary to the expectation that social distance would make cross-caste hiring most costly at the extremes. Combining the taste bias and the hiring wedge, the total barrier facing a firm selling to another caste group averages 21.5% of the marginal cost, ranging from 12.4% (LC selling to HC) to 26.4% (MC selling to LC, Table 6)—comparable to the cross-state trade cost estimates for India in Asturias et al. (2019).

Despite these barriers, there are substantial gains from cross-caste trade (Table 7): full autarky would reduce aggregate real income by 9.7 percent, with LC consumers—the poorest group—bearing the largest loss (11.4 percent). The remaining identity barriers nonetheless impose substantial costs. Decomposing these losses, removing hiring wedges raises aggregate real income by 4.7 percent—roughly twice the 2.6 percent gain from eliminating taste bias.

The gains from removing barriers are distributed unevenly. Removing taste bias barely changes cross-group inequality, as HC consumers also benefit from accessing cheaper LC varieties. Removing hiring wedges, by contrast, reduces inequality by 20%. LC consumers gain the most across all counterfactuals, reflecting their smallest home market and greatest dependence on cross-caste exchange. Asymmetric barriers are particularly regressive: when HC consumers refuse LC varieties entirely ($\beta_{\text{HC,LC}} \to \infty$) while LC consumers continue to buy HC goods ($\beta_{\text{LC,HC}}$ at baseline), aggregate real income is virtually unchanged but inequality rises by 20%. Fully isolating LC from all cross-caste trade reduces aggregate real income by 5.4%, with LC consumers bearing nearly the entire burden (11.4% loss), while inequality rises sharply.

A practical implication of these results is that the cost of policies aimed at reducing identity barriers should not exceed the estimated gains. Moreover, identity-based networks may also generate benefits—for instance, by facilitating intergenerational skill transmission or reducing information barriers (Cassan et al., 2025; Fisman et al., 2017; Munshi, 2019)—so the net effect of removing all caste-based segmentation on real income could be smaller than the gross gains reported here.

A natural concern is that the estimated barriers reflect spatial segregation of caste groups rather than identity-based preferences per se. We extend the benchmark framework to incorporate iceberg-style spatial trading costs between groups. As expected, this reduces the estimated identity-based barriers: counterfactual income gains from removing taste bias and hiring wedges are attenuated in the model with spatial trading costs (Online Appendix C.3). However, it is not obvious that spatial segregation should be treated as exogenous. If residential sorting is itself partly driven by taste-based discrimination, then controlling for spatial distance may partial out part of the very barrier we aim to measure. Disentangling the endogenous and exogenous components of spatial segregation is an important avenue for future research.

Literature Review. First, we add to the literature on misallocation in developing economies. Hsieh and Klenow (2009) show that factor misallocation across firms reduces aggregate TFP substantially in India and China; Boehm and Oberfield (2020) trace misallocation in the market for inputs to relationship-specific frictions between buyers and suppliers.[2] Hsieh et al. (2019) quantify the aggregate cost of occupational barriers facing women and minorities in the U.S. We study a complementary channel: rather than barriers to occupational entry, we focus on hiring frictions and demand-side taste bias that segment product markets along identity lines, and embed them in a quantitative general-equilibrium framework.[3]

Second, we contribute to the growing literature on identity, social networks, and trade. Recent work shows that cultural proximity shapes bilateral trade flows and that within-caste firm-owners are twice as likely to trade as cross-caste firm-owners (Fujiy et al., 2025; Boken et al., 2025). We formalise identity-based barriers as bilateral trade costs between caste-group markets, allowing us to quantify their aggregate cost within a standard trade framework. Our estimates are consistent with existing magnitudes: the trading of new seeds is 50% higher in caste-homogeneous villages (de Janvry et al., 2025), and caste-homogeneous villages have approximately 45% higher agricultural yields due to lower trading barriers (Anderson, 2011).[4]

Third, we contribute to the literature on discrimination. Becker (1957) distinguished customer, employer, and employee channels, and most subsequent work has studied these channels in isolation. On the customer side, discrimination reduces demand for minority entrepreneurs (Borjas and Bronars, 1989).[5] On the employer side, discrimination lowers callback rates for minority job applicants (Kline et al., 2022). On the employee side, discrimination lowers productivity, Hjort (2014).[6]

A recent strand examines the interaction between the two: Holzer and Ihlanfeldt (1998) shows that firms' hiring tracks their customers' demographics, and Cook et al.

[2]Other sources of misallocation in India include capital market barriers (Bau and Matray, 2023) and occupational sorting by caste (Cassan et al., 2025).

[3]Recent work shows that demand-side factors, particularly product appeal, explain most of the variation in firm size (Hottman et al., 2016; Eslava et al., 2024); we identify social identity as a specific mechanism through which appeal is shaped. For related evidence on demand and firm size, see Foster et al. (2008, 2016); Startz (2025); Hardy and Kagy (2020); Bernard et al. (2022); Bold et al. (2022); Bassi et al. (2023); Tan and Zeida (2023).

[4]For related work on networks and trade costs, see Rauch (2001); Rauch and Trindade (2002); Combes et al. (2005); Felbermayr and Toubal (2010); Aker et al. (2014); Desmet et al. (2023). On caste and restricted trade in India, see Banerjee and Munshi (2004) and Emerick (2018).

[5]On demand-side discrimination, see List (2004); Doleac and Stein (2013); Edelman and Luca (2014) for the U.S., Nagavarapu and Sekhri (2016) for India, and Kelley et al. (2024) for Sub-Saharan Africa.

[6]See also Arrow (1973) on statistical discrimination as an alternative mechanism, Charles and Guryan (2008) on prejudice and wages, Hurst et al. (2024) on customer-contact versus abstract tasks, Bohren et al. (2022) on systemic discrimination, and Bertrand and Duflo (2017) for a survey of field experiments on discrimination.

(2023) and Cook et al. (2025) find that firms' willingness to discriminate in hiring mirrors their consumers' biases. Rubinstein (2025) shows—theoretically and using U.S. banking deregulation—that discrimination can be exacerbated by competition. Our paper contributes to this emergent literature. We provide a general-equilibrium model that jointly embeds both channels in the firm's cost structure and allows us to decompose the macroeconomic and distributional consequences of identity-based market segmentation. The key mechanism is destination labour: serving a consumer group requires hiring workers from that group, linking demand-side taste bias and supply-side hiring barriers within the firm.[7]

2 Institutional Background

The Indian caste system is a hereditary social hierarchy characterised by endogamy, hereditary membership, and traditionally prescribed occupations (Béteille, 1965). In its classical form, society was divided into four *varnas*—Brahmins (priests), Kshatriyas (warriors), Vaishyas (merchants), and Shudras (labourers)—with Scheduled Castes (Dalits) and Scheduled Tribes (Adivasis) historically placed outside the ritual hierarchy. The hierarchy was enforced through practices of social exclusion: higher-ranked castes restricted physical contact and commensality with lower-ranked groups, and assigned them stigmatised occupations such as manual scavenging and leather work (Dumont, 1980). Despite the constitutional prohibition of untouchability after independence and extensive affirmative action through the reservation system, caste remains the primary marker of social identity in India, continuing to govern marriage patterns and everyday economic interactions (Srinivas, 1994; Munshi, 2019). Because caste is endogamous—one is born into a caste and cannot leave it—identity is permanent across generations, distinguishing it sharply from income or class.

Modern Indian administrative systems classify households into three broad groups, which we adopt throughout: High-ranked Castes (HC, corresponding to the General category), Middle-ranked Castes (MC, corresponding to Other Backward Classes), and Low-ranked Castes (LC, corresponding to Scheduled Castes and Scheduled Tribes, who share similar economic disadvantages in the data despite distinct historical origins). Each of these categories subsumes thousands of endogamous sub-groups known as *jatis*, which are the operative units of social identity in daily life. Because our data classify households at the broad category level, our estimates of identity-based barriers reflect *between*-category preferences and should be interpreted as lower bounds on

[7]On caste and economic outcomes more broadly, see the survey by Munshi (2019), and Oh (2023) on caste and occupational choice, Madheswaran and Attewell (2007); Hnatkovska et al. (2012) on labour markets, Bhagavatula et al. (2023) on caste diversity in boardrooms, and on gender norms Agte and Bernhardt (2023).

the barriers that would obtain at the jati level, where identity is more finely observed and social distance more precisely calibrated (Boken et al., 2025; Fujiy et al., 2025).

The caste system has historically functioned as a division of labour, institutionalised through the *jajmani* system, in which each caste group performed specialised services for other castes—Brahmins officiated religious ceremonies, barbers provided grooming, potters supplied earthenware—receiving grain or other goods in return (Wiser, 1936; Kolenda, 1963). For such occupationally specialised goods, the relevant economic relationship is one of complementarity, not within-group preference, and cross-caste exchange was essential to village self-sufficiency.

Our mechanism operates in a different domain: the substantial and growing segment of the economy where multiple caste groups produce and sell the same goods. As India's economy has transformed, the traditional occupational assignments of the jajmani system have weakened considerably (Béteille, 1965; Munshi and Rosenzweig, 2006). Industrialisation rendered many hereditary occupations redundant, new occupations emerged outside the caste framework, and the incorporation of villages into market economies loosened inter-caste ties (Gould, 1964; Srinivas, 2003). It is within these overlapping sectors—where consumers can choose between caste-proximate and caste-distant producers of the same good—that identity-based preferences create demand segmentation. Our empirical strategy focuses on these overlapping sectors precisely to distinguish identity preferences from occupational specialisation.

Table A.4 summarises the economic characteristics of the three groups using NSS household data. In rural India, LC households account for 33% of the population. Their monthly per-capita expenditure is 68% of the HC level; MC households fall in between at 83%. The literacy rate among LC household heads is 43%, compared with 55% for MC and 70% for HC. Among LC workers, 59% are employed in agriculture, compared with 47% of MC and 26% of HC workers. Two features of this income gradient matter for our analysis. First, because LC households are closer to subsistence, their consumption is more sensitive to income shocks—a property we exploit when using rainfall variation to identify demand-side taste bias. Second, the correlation between income and caste identity means that any observed relationship between a firm's revenue and the caste composition of its local market could reflect either identity-based preferences or differences in purchasing power. The rainfall instrument isolates exogenous shifts in caste-specific purchasing power, and the model's non-homothetic structure allows income differences to operate through market size while the taste parameter β captures the residual demand segmentation that income alone cannot explain.

Our mechanism requires that consumers observe, or can infer, the caste identity of the producer. This condition is plausible in rural MSME markets, where transactions are face-to-face and the firm owner is a known member of the community. Caste is transmitted through surnames, dialect, and social networks (Srinivas, 1962). In sectors such as tailoring, furniture and woodwork, and repair and maintenance work, the consumer interacts directly with the owner or employees, making social identity immediately salient. Lowe (2021) documents significant caste-based homophily in face-to-face interactions, with individuals twice as likely to form connections with someone from the same caste. By contrast, when consumers purchase goods from large formal firms, they neither know nor plausibly care about the caste composition of the workforce. Identity-based demand preferences are therefore most likely to bind in the local, non-tradable markets that constitute the bulk of rural activity. One domain where these preferences are especially acute is food: the caste hierarchy is partly defined by rules governing who may prepare and handle food, with higher-ranked castes traditionally refusing food from lower-ranked groups (Dumont, 1980). We exploit this additional layer of caste salience in our empirical strategy (Fact 3, Section 4.3).

3 Data and Measurement

We use three data sources: the 2006–07 MSME Census for firm-level outcomes, the National Sample Survey for household consumption, and satellite rainfall data for identification.

3.1 Micro, Small, and Medium Enterprise (MSME) Data

India's micro, small, and medium enterprise (MSME) sector is the second-largest source of employment after agriculture, with approximately 30% of GDP, 35% of manufacturing output, and 46% of total exports (Ministry of Micro, Small and Medium Enterprises, 2024). In rural areas—the focus of our analysis—MSMEs are the primary source of non-agricultural employment and income. We use the 2006–07 MSME Census, a nationally representative survey that uniquely records the caste of both the enterprise owner and employees—information missing in other firm-level datasets such as the ASI and Prowess. Unlike ASI and Prowess, which cover large firms, MSME focuses on enterprises below a capital threshold. This is appropriate for our analysis, as single-establishment firms operating in local markets are precisely the setting where identity-based demand barriers are most relevant.[8]

A key feature of this sector is that production is not segregated by caste. Online Appendix Table A.1 classifies sectors by caste concentration: at the 4-digit NIC level, overlapping sectors—those where no single caste group owns more than 80% of firms—account for 177 of 195 sectors, 99% of firms, and 98% of revenue. Online

[8] In the model, we also include a homogeneous goods sector, which is not subject to consumer bias. Large firms represent a relatively small fraction of total firms in India (Hsieh and Olken, 2014).

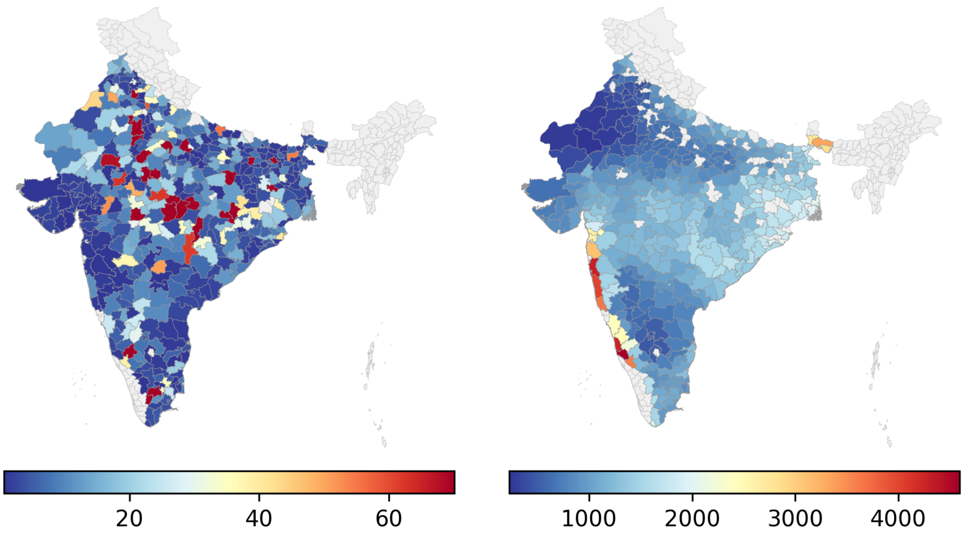
Table 1: Firm-size distribution

	All Firms						Mean by Owner Caste		
	Mean	Median	p5	p95	N	Firms (wt.)	HC (191,084)	MC (197,634)	LC (60,587)
Emp. All	5.5	2	1	15	444,580	449,305	7.8	4.0	3.4
Emp. LC	1.2	0	0	4	444,580	449,305	1.6	0.6	1.7
Emp. MC	2.0	1	0	6	444,580	449,305	1.9	2.6	0.7
Emp. HC	2.3	1	0	7	444,580	449,305	4.3	0.8	1.0
Emp. Own-C (%)	74	100	0	100	444,574	449,305	69	80	68
Revenue (10^4)	366	13.5	2.9	580	444,580	449,305	663	129	208
Materials (10^4)	243	4.4	0.2	349	444,580	449,305	453	75	130

Notes. The table presents the firm size distribution of rural registered MSME firms. 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. N is the unweighted number of firms in the sample; Firms (wt.) is the weighted count using MSME sampling multipliers. The last three columns report the mean for firms owned by each caste group; the number of weighted firms in each group is shown in parentheses below the column header. Source: MSME 2006–07.

Appendix Table A.2 lists the ten largest overlapping sectors by revenue share. In grain milling (13.3% of MSME revenue), HC entrepreneurs own 49% of firms, MC own 43%, and LC own 8%. In textiles, wearing apparel, and metalwork, all three groups are active producers, often within the same district. This pervasive overlap means that consumers routinely face a choice between caste-proximate and caste-distant producers of the same good—the competitive environment in which identity-based demand preferences can shape firm outcomes. Given that the vast majority of economic activity occurs in overlapping sectors, and that our main patterns hold equally in both overlapping and caste-specialized sectors (see Fact 1), we use the full sample for all subsequent analyses.

The MSME survey consists of registered and unregistered firms as defined under the Factories Act 1948. We focus on registered firms as they are 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 focus on firms in rural areas, where caste norms are stronger, local markets are more segmented, and rainfall shocks provide a clean source of exogenous variation in caste-specific demand through their effect on agricultural wages. After cleaning, we have 444,574 firms employing approximately 3 million employees in 195 sectors producing 4,590 distinct products and services; see Online Appendix A.2 for more details. We provide the distribution of employees and revenues in Table 1. On average, LC-owned firms are present across India but they are small. Figure 1 plots the difference in revenues between HC- and LC-owned firms, highlighting substantial variation across districts that we will exploit in our empirical analysis.



(a) HC/LC-owned firms (b) Rainfall

Figure 1: Spatial Variation in Rainfall and Firm Size across Districts

Notes. Figure (a) uses MSME 2006-07 data to plot the absolute difference in log(gross output value) between HC- and LC-owned firms. Figure (b) plots the distribution of rainfall across districts.

3.1.0.1 Balanced panel. The MSME data provides retrospective information on *revenues* and *material purchased* for the firms that survive up to 2006-2007. This allows us to construct a balanced panel of MSMEs for the three years, 2004-05 to 2006-07.

3.1.0.2 Prices and quantities. The MSME data provides 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. We use this data to distinguish between the price effect and the quantity effect of demand shocks.

3.2 National Sample Survey (NSS) Household Data

We use household-level data from the National Sample Survey (NSS) Household Consumer Expenditure Survey (Schedule 1.0), which records detailed consumption expenditure by caste across product categories. We use data from multiple survey rounds spanning 2004–2007 to construct a district-level panel. Summary statistics are provided in Table A.4. We also use NSS Employment and Unemployment Survey data to measure wage elasticity; details are provided in Online Appendix A.

3.3 Rainfall Data

We use rainfall 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 1b plots the spatial variation in rainfall across districts in India during 2006-07, measured in millimetres, showing substantial cross-district variation. For the analysis, we follow Jayachandran (2006) and define rainfall shock as equal to +1 for a positive shock (historically good rainfall), -1 for a negative shock (historically bad rainfall), and 0 otherwise.

All analyses use sampling multipliers provided in the MSME census, and standard errors are clustered at the district level unless otherwise noted.

4 Stylized Facts

We document four stylized facts on cross-caste hiring patterns and their relationship to consumer demand. Fact 1 establishes that cross-caste hiring is widespread but asymmetric. Fact 2 shows that larger firms hire more diverse workforces, especially in contact-intensive sectors where employee–customer interaction is prevalent. Fact 3 demonstrates that the caste composition of local consumer demand correlates strongly with firm employment composition. Fact 4 exploits rainfall shocks as exogenous variation in caste-specific demand to establish causal evidence.

4.1 Fact 1: Cross-caste Hiring is Prevalent but Asymmetric

Table 2 presents the workforce composition and hiring participation patterns across caste groups. A random benchmark implies that the share of workers should be HC: 42.0 / MC: 36.7 / LC: 21.3—share of workers of each caste in the whole sample. This reflects overall employment shares in the data. Panel A reports the actual hiring patterns at the intensive margin: the share of each caste group in total employment, by owner caste. The diagonal entries—own-caste workers—dominate: 49.7% of employees in LC-owned firms are LC, 64.2% in MC-owned firms are MC, and 54.8% in HC-owned firms are HC. Strong in-group homophily in hiring is the norm.

Yet the off-diagonal entries reveal substantial cross-caste hiring. About 50.4% of LC-owned firms' workforce belongs to other castes, compared to 35.8% for MC-owned and 45.1% for HC-owned firms. This cross-caste hiring is asymmetric. HC workers account for 30% of employment in LC-owned firms, while LC workers account for only 20.5% of employment in HC-owned firms. Similarly, MC-owned firms employ few LC workers (15.4%), whereas LC-owned firms employ 20.3% MC workers.

Panel B shifts attention to the extensive margin (hiring participation). Each cell reports the share of firms that employ at least one worker of a given caste. Off the diagonal, the patterns are strikingly asymmetric. Only 14% of MC-owned firms hire any LC workers, compared to 24% of HC-owned firms. Across all firms, 38% hire at least one out-caste worker. These patterns are robust to alternative sample restrictions (urban vs. rural areas, overlapping sectors) as shown in Online Appendix Table A.7.

Table 2: Cross-Caste Hiring Patterns

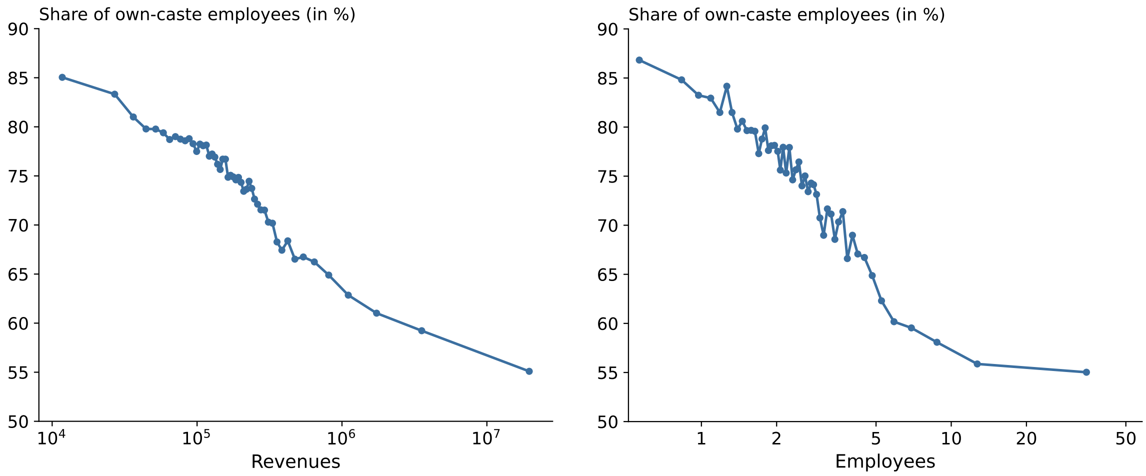
	Employment Shares (%)			Extensive Margin (%)		
	HC-owned	MC-owned	LC-owned	HC-owned	MC-owned	LC-owned
Panel A: Rural, all sectors						
HC workers	54.8	20.4	30.1	87.6	21.9	25.6
MC workers	24.6	64.2	20.3	38.7	89.5	18.7
LC workers	20.5	15.4	49.7	23.5	13.6	76.1
Random benchmark	HC: 42.0 / MC: 36.7 / LC: 21.3					
Panel B: Rural, overlapping sectors						
HC workers	54.8	20.4	30.0	87.5	21.9	25.6
MC workers	24.7	64.3	20.3	38.8	89.5	18.6
LC workers	20.5	15.3	49.7	23.4	13.6	76.1
Random benchmark	HC: 41.8 / MC: 36.8 / LC: 21.3					

Notes. Panel A (Employment Shares) reports the share of total employment by worker caste (rows) within firms grouped by owner caste (columns). Panel B (Extensive Margin) reports the share of firms (by owner caste) that employ at least one worker of the indicated caste. Sample: rural registered MSME firms. Source: MSME 2006–07. Sampling multipliers are applied.

4.2 Fact 2: Larger Firms Have Diverse Workforces

Figure 2 shows a steep, monotonic decline in the own-caste share as firms grow—from about 85% among the smallest to roughly 58% among the largest—and the pattern holds across alternative size measures (Figures 2a and 2b). This negative relationship is robust in urban areas (Online Appendix Figure A.1) and holds when measuring workforce diversity using the caste Herfindahl–Hirschman Index (Online Appendix Figure A.2). Detailed decomposition by owner caste (Online Appendix Figure A.3) confirms that larger firms of all castes hire more diverse workforces.

The relationship between firm size and workforce diversity is stronger in customer-facing sectors, where interaction between employees and consumers is salient. This result is consistent with recent literature indicating that customer discrimination is prevalent in these sectors; see Kline et al. (2022) and Rubinstein (2025). We define contact-intensive sectors as industries where customer–employee interaction is prevalent: carpentry and furniture manufacturing, construction, wholesale and retail trade, hotels and restaurants, travel agencies, post and telecommunications, and computer-related services. Online Appendix Table A.13 lists the detailed NIC-2 industry classifi-



(a) Firm size measured by revenue

(b) Firm size measured by employment

Figure 2: Homophily in Hiring and Firm Size

Notes. Figure 2a 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. Figure 2b presents a bin-scatter plot; the x-axis is the total employees, and the y-axis is the share of Own-caste workers. Sampling multipliers are applied.

cations and reports the distribution of firms by owner caste. In these contact-intensive industries, the slope is an additional -3.8 pp steeper for LC-owned firms (Online Appendix Table A.8, Column 4). This pattern is consistent with demand-driven diversification: LC-owned firms face tighter in-group demand—their peer consumers are poorer—and expand by hiring workers from other castes to reach broader consumer markets. Pure supply-side explanations (e.g., difficulty in recruiting own-caste workers as firms grow) fit the data less well, since the size–diversity gradient is markedly stronger in customer-facing sectors where employee–consumer interaction matters.

4.3 Fact 3: Employee Shares and Consumption Shares

We now examine the cross-sectional relationship between the caste composition of local consumer demand and the caste composition of firm employment. A key identification challenge is that districts with more LC residents mechanically have both more LC consumers and more LC workers, making it difficult to separate demand-side effects from labour supply composition. To address this concern, we exploit variation across product categories: if demand preferences drive employment patterns, the correlation should be stronger for products where caste-based preferences are most intense (food products, where purity norms are salient) compared to other products, even though firms in the same district draw from the same labor pool.

We match MSME product codes to NSS consumption categories at 16 product types (e.g., cereals & pulses, dairy/oil/sugar, clothing & bedding, services), which roll up into 8 broad groups (food, tobacco, fuel, clothing, footwear, medical, miscellaneous, and durables). The concordance and firm distribution by caste across product categories are provided in Online Appendix Table A.5. Using this merged dataset, we estimate the following regression for each pair of caste groups (s, s'):

$$\text{EmpShare}_f^{s'} = \gamma_{s \to s'} \cdot \text{DemandShare}_{dp}^{s} + \alpha_r + \alpha_p + \varepsilon_f, \tag{1}$$

where $\text{EmpShare}_f^{s'}$ is the share of caste-s' workers in firm f's total employment, and $\text{DemandShare}_{dp}^{s}$ is the share of caste-s consumption expenditure in total household spending for product category p in district d, computed from NSS data. We include state fixed effects (α_r) and product fixed effects (α_p), and cluster standard errors at the district level. Each coefficient $\gamma_{s \to s'}$ captures whether districts where caste s accounts for a larger share of consumer demand for a given product also have firms with a higher share of caste-s' employees. We estimate all nine pairwise combinations ($s, s' \in \{\text{HC}, \text{MC}, \text{LC}\}$) and present the results as 3×3 coefficient matrices.

Panel A of Figure 3 presents the 3×3 matrix of $\hat{\gamma}_{s \to s'}$ estimates from Equation (1). The diagonal entries are all positive and significant: $\hat{\gamma}_{\text{HC}\to\text{HC}} = 0.108$ ($p < 0.01$), $\hat{\gamma}_{\text{LC}\to\text{LC}} = 0.079$ ($p < 0.01$), and $\hat{\gamma}_{\text{MC}\to\text{MC}} = 0.120$ ($p < 0.01$). This means that in districts where a given caste group accounts for a larger share of consumer demand, firms employ more workers of that same caste group. The off-diagonal entries are generally negative, indicating a crowding-out pattern: higher own-caste demand is associated with lower employment of other-caste workers.

To implement this identification strategy, we estimate a second regression that augments Equation (1) by interacting the demand share with a food-product indicator:

$$\begin{aligned} \text{EmpShare}_f^{s'} = {} & \gamma_{s \to s'} \cdot \text{DemandShare}_{dp}^{s} + \delta_{s \to s'} \cdot \text{DemandShare}_{dp}^{s} \times \text{Food}_p \\ & + \alpha_r + \alpha_p + \varepsilon_f. \end{aligned} \tag{2}$$

The food interaction diagonal is positive and significant: food firms respond more strongly to local demand composition than non-food firms. Since food and non-food firms within the same district draw from the same labour pool, this amplification cannot be explained by labour supply composition and instead reflects the demand-side identity channel.

Panel B of Figure 3 reports the $\hat{\delta}_{s \to s'}$ coefficients from this regression. The diagonal entries are positive—HC: 0.074 ($p < 0.10$), LC: 0.092 ($p < 0.01$), MC: 0.079 ($p < 0.05$)—confirming that the demand–employment linkage is significantly amplified for food products across all three caste groups. Online Appendix Table A.9 provides extensive robustness: it replicates the analysis for urban samples, splits food and non-food products separately (showing diagonal coefficients are roughly twice as large for food), and adds district-level controls (urbanization, agricultural share, literacy rate, and caste residential segregation) to absorb confounding local characteristics.

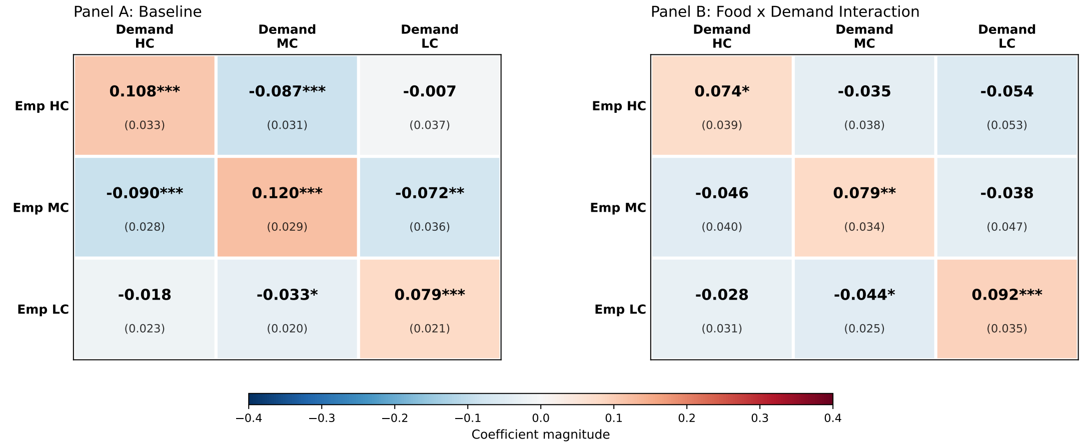


Figure 3: Cross-Caste Demand–Employment Matrices: Baseline and Food Interaction

Notes. Panel A reports $\hat{\gamma}_{s \to s'}$ from Equation (1) (baseline, without food interaction). Panel B reports the food interaction coefficients $\hat{\delta}_{s \to s'}$ from the separate regression in Equation (2). The two panels come from different regressions. Rows index the caste of employees; columns index the caste whose demand share is the regressor. Darker red (blue) indicates larger positive (negative) coefficients. Coefficients and standard errors (in parentheses) are displayed in each cell. All regressions include state and product fixed effects ($N = 333{,}639$). Standard errors are clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.4 Fact 4: Consumption Shocks, Firm Revenues, and Hiring

The previous facts establish correlations between demand composition and employment patterns, but correlation does not imply causation. To identify the causal effect of demand-side preferences on firm outcomes, we exploit rainfall shocks as an exogenous source of variation in caste-specific demand. Since LC households are disproportionately employed in agriculture, positive rainfall shocks raise their wages and consumption expenditure relative to other caste groups. If demand-side identity preferences matter, we should observe that: (i) LC household consumption responds more strongly to rainfall; (ii) this demand shift asymmetrically benefits firms owned by or employing LC workers; and (iii) HC-owned firms respond by hiring more LC workers to capture the expanded LC consumer market. We test each prediction in turn.

4.4.1 Household consumption elasticity

To test this prediction, we estimate the following equation:

$$\log(c_{ht}) = \alpha + \rho_1 \cdot \text{Rainshock}_{dt} + \theta_s \cdot \text{Rainshock}_{dt} \times \text{caste}_s + \delta_{d,t} + \delta_s + \epsilon_{ht}, \quad (3)$$

where h indexes households, d districts, t years, and s caste groups. We include caste fixed effects to absorb any time-invariant caste traits, and district-year fixed effects to capture district-specific shocks. The coefficient θ_{LC} gives the sensitivity of LC households' consumption to a positive income shock due to rainfall. If LC households are more dependent on agricultural income, we expect $\theta_{\text{LC}} > 0$.

Table 3 presents the results from Equation (3), and shows the asymmetric effect of rainfall on Monthly Per Capita Expenditure (MPCE). LC households' consumption increases by 9.2% in districts with a positive rainfall shock relative to HC households. We also find an increase in the MC households' MPCE. Importantly, the increase in consumption is disproportionately concentrated in non-food categories. While spending on food and non-food items combined increases by 7.4% (Column 2), spending on services and durables increases by 15.0% and 21.8% respectively (Columns 5–6), indicating substantial consumption upgrading beyond basic necessities. This reflects the usual Engel curve pattern; income shocks shift expenditure shares away from necessity goods. MC households' demand also increases, but the elasticity is smaller relative to LC households.[9] The consumption response is symmetric for positive and negative rainfall shocks (Online Appendix Table A.10), ruling out one-tailed effects. The income channel operates primarily through agricultural wages: Online Appendix Table A.11 shows LC agricultural wages rise by 5.7% following positive rainfall, with no response for non-agricultural wages.

4.4.2 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 firm outcomes. We regress the log of gross output (revenue) of firm i in district d in year t on district-level rainfall shocks interacted with the caste group of the firm owner:

$$\ln y_{idt} = \rho_s \cdot \text{Rshock}_{dt} \times \mathbf{1}[\text{caste} = s] + \alpha_{dt} + \alpha_{pt} + \gamma_s + \varepsilon_{idt}, \tag{4}$$

where Rshock_{dt} is the rainfall shock as defined in Section 3.3. All regressions include district $\times$ year fixed effects (α_{dt}) and product code $\times$ year fixed effects (α_{pt}, using all 4,590 MSME product codes), and are weighted by survey multipliers. Standard errors are clustered at the district level. The coefficients ρ_{MC} and ρ_{LC} measure the revenue elasticity of MC- and LC-owned firms with respect to the rainfall shock, relative to HC-owned firms.

Figure 4 summarizes the LC revenue elasticity across all specifications with 90%

[9] In a series of robustness checks, we find LC households' consumption response to be much more robust than the MC households', see Online Appendix A.

Table 3: Consumption Elasticity

	log MPCE	log Food +Non-food	log Fuel	log Clothing	log Edu+Med +Services	log Durables
Rainshock × MC	0.045**	0.036*	0.026	0.050**	0.080**	0.115*
	(0.021)	(0.022)	(0.024)	(0.025)	(0.036)	(0.067)
Rainshock × LC	0.092***	0.074***	0.046*	0.108***	0.150***	0.218***
	(0.028)	(0.026)	(0.027)	(0.032)	(0.044)	(0.072)
Observations	111,312	111,266	111,171	111,191	99,436	97,024
R-squared	0.275	0.266	0.338	0.258	0.306	0.195
Exp. share (MC)	1	0.620	0.087	0.065	0.218	0.033
Exp. share (LC)	1	0.643	0.094	0.065	0.173	0.026
Exp. share (Avg)	1	0.613	0.088	0.064	0.212	0.033
Controls	✓	✓	✓	✓	✓	✓
Caste FE	✓	✓	✓	✓	✓	✓
District × Year FE	✓	✓	✓	✓	✓	✓

Notes. This table reports the elasticity of household consumption to rainfall shocks by caste group. Column (1) uses log monthly per-capita expenditure (MPCE). Columns (2)–(6) decompose consumption into: food and non-food items combined, fuel & light, clothing & footwear, education and medical services, and durables. Education and institutional medical are recorded on a 365-day recall (NSS Block 9); miscellaneous services including non-institutional medical on a 30-day recall (Block 10); all converted to monthly equivalents. Expenditure shares are reported as the ratio of aggregate spending for MC, LC, and overall. Controls include the household head's number of meals per day, education level, and land ownership, each interacted with the rainfall shock. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

confidence intervals, confirming that the effect is positive and significant throughout. Table A.12 presents the full results from Equation (4). In the baseline specification without firm-level controls (column 2), a positive rainfall shock raises the revenues of LC-owned firms by approximately 11.1% relative to HC-owned firms. Adding firm-level controls—firm age, revenue quartile, and firm characteristics interacted with caste (column 3)—the LC coefficient attenuates to 4.7% but remains highly significant ($p < 0.01$). Columns (4)–(13) show that this result is robust across a range of subsample splits: high consumption elasticity sectors, non-tradable goods, products using similar quality intermediates across castes, unsegregated districts, large firms, and sectors not dominated by LC firms. We will use the lowest estimate of 3.6% as our reference number for calibration purposes in Section 6.1.2.

Detailed discussion in Appendix A.6.5 tests alternative mechanisms, including product market segmentation, price discrimination (output and input prices), spatial segregation, and firm characteristics, finding no evidence that these confounds drive the results.

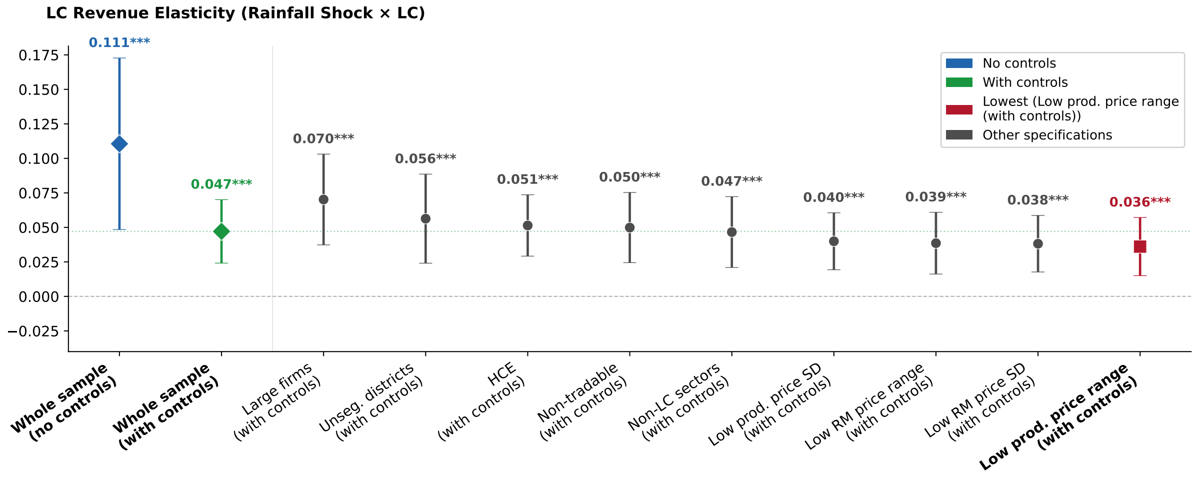


Figure 4: LC Revenue Elasticity Across Specifications

Notes: Each point reports the coefficient on Rainfall Shock × LC from the corresponding column of Table A.12, with 90% confidence intervals. "Whole sample" is the benchmark specification with product × year and district × year FE, firm-level controls, and revenue quartile FE (column 3). The remaining specifications restrict the sample to subgroups as labeled, sorted in decreasing order of the LC elasticity. The lowest elasticity is highlighted in red.

4.4.3 Foreign Demand Shocks and Firm Outcomes

A potential alternative explanation for the asymmetric revenue effects is that LC workers are simply cheaper, and expanding firms hire them to reduce labor costs regardless of consumer identity. This supply-side channel would imply that any demand shock—not just caste-specific ones—should lead firms to hire more LC workers. To test this, we conduct a placebo exercise exploiting foreign demand shocks, which expand firm revenues but are naturally caste-neutral. 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 (see Online Appendix A.7). We find that foreign demand shocks are indeed expansionary; however, they do not lead to an asymmetric increase in demand for LC employees (see Table A.16). This rules out the cost-based explanation and supports the demand-side identity channel.

4.4.4 Workforce Composition

The revenue elasticity results show that LC-owned firms benefit from rainfall-induced LC demand shocks, but do HC-owned firms also respond by adjusting their workforce composition to capture this demand? To test the demand channel more sharply, we examine HC-owned firms' hiring responses in two settings where demand considerations should be especially salient: (1) competitive markets, where LC consumers have credible outside options and can switch to LC-owned competitors if not served by HC firms, creating stronger incentives for HC firms to hire LC workers as a demand-capturing strategy; and (2) contact-intensive sectors, where employee–customer interaction makes caste identity more salient and targeted hiring more valuable. If demand-side preferences drive hiring, HC firms should hire more LC workers following positive LC demand shocks in both settings.

We exploit variation in market competition and sectoral heterogeneity to identify the response of HC-owned firms. Within HC-owned firms, we estimate:

$$y_f = \alpha + \rho_1 \cdot \text{Rainshock}_d + \theta \cdot \text{Rainshock}_d \times X_{d,j} + \delta_d + \delta_j + \epsilon_f, \tag{5}$$

where y_f is the share of LC employees, δ_d and δ_j denote district and NIC-4 industry fixed effects, respectively, and $X_{d,j}$ is either (i) $Competitive_{d,j}$, an indicator for whether the HC share of firms in a district $\times$ product market falls below a given threshold, or (ii) $Contact\text{-}Intensive_j$, an indicator for sectors where customer–employee interaction is prevalent. In markets where HC firms dominate, the incentive to hire LC workers is low, as LC consumers have few outside options. Similarly, in contact-intensive sectors, caste identity becomes more salient. We expect $\theta > 0$ in both cases.

4.4.4.1 Workforce composition of HC firms in competitive markets. We estimate Equation (5) within HC firms. We define a market as competitive if the HC share of firms in a district $\times$ product code market falls below 50%. Figure A.4 presents the interaction coefficient across alternative definitions of competitiveness, and Table 4 (Columns 1–2) reports the baseline specification. On average, HC firms in competitive markets have a lower LC employee share than those in HC-dominated markets (Competition coefficient of -0.024, Column 1). However, the interaction of the rainfall shock with competition is positive and significant (0.016, $p < 0.05$), indicating that when LC household demand rises, HC firms in competitive markets hire relatively more LC employees compared to those in HC-dominated markets. This effect is robust to restricting the sample to high consumption elasticity sectors (Column 2). These results indicate that competition drives demand-led diversification, as firms capture consumer demand via targeted hiring.

4.4.4.2 Workforce composition of HC firms in contact-intensive sectors. In our final strand of evidence on workforce composition, we run the regression in Equation (5) within HC-owned firms and provide results in Table 4 (Columns 3–4). The interaction of the rainfall shock with the contact-intensive indicator is positive and significant: HC firms in contact-intensive sectors hire 0.010 ($p < 0.05$) more LC employees when local LC demand rises (Column 3). The effect is larger when restricted to high consumption elasticity sectors (0.013, $p < 0.05$, Column 4). This is in contrast to the baseline cross-sectional pattern, where LC employee share in contact-intensive sectors is lower on average, consistent with caste identity becoming more salient when there

Table 4: LC Employee Share Among HC-owned Firms: Competition and Contact-Intensive Industries

	Dep. var.: Share of LC workers			
	Competition		Contact-intensive	
	(1) All	(2) HCE	(3) All	(4) HCE
Competition	-0.024***	-0.015***		
	(0.004)	(0.004)		
Rshock × Competition	0.016**	0.016**		
	(0.008)	(0.008)		
Rshock × Contact			0.010**	0.013**
			(0.005)	(0.005)
District FE	✓	✓	✓	✓
NIC-4 FE	✓	✓	✓	✓
Observations	179,958	130,953	179,958	130,953
R^2	0.184	0.203	0.182	0.203
Mean LC share	0.096	0.095	0.096	0.095

Notes. The dependent variable is the share of LC (SC/ST) employees in total firm employment among HC-owned firms. *Competition* is an indicator equal to one if the HC share of firms in the district × product code market is below 50 percent. *Rshock* is the 2006 rainfall shock at the district level. *Contact* is an indicator for contact-intensive industries, defined as sectors where customer–employee interaction is prevalent: carpentry and furniture manufacturing, construction, wholesale and retail trade, hotels and restaurants, travel agencies, post and telecommunications, and computer-related services. *HCE* denotes high consumption elasticity sectors, defined as all sectors excluding food, non-food consumer goods, and fuel. Columns (1)–(2) report the interaction of the rainfall shock with the competition indicator; Columns (3)–(4) report the interaction of the rainfall shock with the contact-intensive indicator. Columns (2) and (4) restrict the sample to HCE sectors. All regressions include district and NIC-4 industry fixed effects. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

is customer–employee interaction.

A limitation of our workforce composition analysis is that caste-specific employment is observed only in the 2006–07 cross-section, so we cannot estimate within-firm changes in hiring over time in response to rainfall shocks. However, our identification strategy does not require panel variation in workforce composition. All specifications in Equation (5) include district fixed effects, which absorb any district-level confound—including rainfall-induced shifts in LC labour supply, LC wages, or the overall availability of LC workers. Identification comes from the *interaction* of the rainfall shock with product-market competition or contact-intensity, which varies across sectors and product markets *within the same district*. A district-level change in LC labour supply has no reason to differentially affect competitive versus non-competitive product markets, or contact-intensive versus non-contact-intensive sectors, within the same district. The only explanation consistent with the interaction pattern is demand-driven: in competitive markets, HC firms face pressure from LC-owned rivals and must hire LC workers to retain LC consumers; in HC-dominated markets, this incentive is absent. The cross-sectional design therefore isolates the demand channel through within-district, across-market variation that supply-side alternatives cannot explain.

Online Appendix A.6.6 provides extensive robustness: Figure A.4 tests alternative thresholds for market competition (50%, 60%, 70%, 80%) and market definitions (product code vs. NIC-4 industry); Tables A.13 and A.14 examine alternative definitions of contact-intensive sectors (baseline vs. broad); and Table A.15 confirms the results hold with firm-level controls.

4.4.4.3 Why indirect evidence is sufficient. Ideally, one would directly link household purchases to the caste composition of the firms they buy from. Such matched consumer–firm transaction data does not exist for India. Nevertheless, the four empirical results in this section collectively identify the demand channel without it. First, LC-owned firms' revenues respond more strongly than HC-owned firms' to rainfall-induced increases in LC demand (Section 4.4), establishing that revenue flows along caste lines. Second, HC-owned firms hire more LC workers in competitive markets—where LC consumers have credible outside options—but not in HC-dominated markets (Section 4.4.4), which is consistent with demand-driven diversification but not with supply-side cost minimisation. Third, the hiring response is concentrated in contact-intensive sectors where employee–consumer interaction is salient, ruling out explanations that are orthogonal to consumer observability of caste. Fourth, foreign demand shocks—which are expansionary but caste-neutral—produce no differential hiring of LC workers (Section A.7), confirming that the asymmetric response is specific to caste-differentiated local demand rather than a mechanical feature of firm expansion. No single alternative—labour-supply shifts, financial frictions, or occupational sorting—can simultaneously account for all four patterns.

5 Theory

We develop a heterogeneous-firm trade model adapted to India's caste system.

5.1 Environment

There are $\mathcal{S} = 3$ groups (castes). Each group s has a labour endowment L_s. Labour is immobile across groups. Wages are $\{w_s\}_{s\in\mathcal{S}}$.

5.1.1 Households

The representative household of group s has non-homothetic preferences with a subsistence requirement $\bar{h}$ for the homogeneous good:

$$
\begin{aligned}
\mathcal{U}(C_{H,s}, C_{D,s}) &= \max_{C_{H,s}, \{c(z,s,s')\}} \; a \log(C_{H,s} - \bar{h}) + (1-a) \log C_{D,s} \\
C_{D,s} &= \left[\sum_{s'} \int_{\Omega_{s'}} \Psi_{s,s'}^{1/\sigma} \, c(z,s,s')^{\frac{\sigma-1}{\sigma}} \, dz \right]^{\frac{\sigma}{\sigma-1}} \\
I_s = w_s L_s + \Pi_s &\geq P_H C_{H,s} + \sum_{s'} \int_{\Omega_{s'}} p(z,s,s') \, c(z,s,s') \, dz,
\end{aligned}
\tag{6}
$$

where $C_{H,s}$ is homogeneous-good consumption, $\bar{h}$ is the subsistence requirement (identical across groups), $C_{D,s}$ is the CES differentiated-good bundle, and $\Omega_{s'}$ is the set of varieties produced by group-s' firms. Non-homotheticity implies that the share of income spent on differentiated goods increases with per-capita income. Poorer groups allocate a larger fraction of income to the subsistence good, leaving less for differentiated varieties.

The taste shifter $\Psi_{s,s'}$ captures demand-side identity bias:

$$
\Psi_{s,s'} = e^{-\beta_{s,s'}}, \qquad \text{with } \beta_{s,s} = 0 \text{ and } \beta_{s,s'} \geq 0 \text{ for } s \neq s', \tag{7}
$$

where $\beta_{s,s'}$ is the taste-for-identity parameter. In general, $\beta_{s,s'}$ is **asymmetric**: $\beta_{s,s'} \neq \beta_{s',s}$. For instance, HC consumers may discriminate more strongly against LC goods ($\beta_{\text{HC,LC}}$ high) than LC consumers discriminate against HC goods ($\beta_{\text{LC,HC}}$ low). In our calibration, we impose a common $\beta_{s,s'} = \beta$ for all $s \neq s'$; differential market access then arises from the asymmetric hiring wedges g_k and export costs c_x. The first-order conditions yield expenditure shares:

$$
P_H C_{H,s} = a(I_s - P_H \bar{h}) + P_H \bar{h}, \quad P_{D,s} C_{D,s} = (1-a)(I_s - P_H \bar{h}). \tag{8}
$$

Expenditure on differentiated goods is a fraction $(1-a)$ of income above subsistence, $I_s - P_H \bar{h}$. The demand for each variety is $c(z,s,s') = \Psi_{s,s'} \, p(z,s,s')^{-\sigma} \, P_{D,s}^{\sigma-1} \, (1-a)(I_s - P_H \bar{h})$, and the market demand index is:

$$
\kappa_s = (1-a)(I_s - P_H \bar{h}) \, P_{D,s}^{\sigma-1}. \tag{9}
$$

Since the homogeneous good serves as the numeraire ($P_H = 1$, see below), subsequent equations write $\kappa_s = (1-a)(I_s - \bar{h}) \, P_{D,s}^{\sigma-1}$.

Non-homotheticity and inequality. Even with equalised nominal wages ($w_s = 1$, see below), groups differ in total income through L_s. A group with lower $(I_s - \bar{h})$ constitutes a smaller market (κ_s lower), creating an amplification mechanism: a smaller market attracts fewer varieties $\rightarrow$ higher price index $\rightarrow$ lower real income.

5.1.2 Production

Homogeneous good. Produced under CRS with one unit of own-group labour per unit of output. Freely traded across groups, it serves as the numeraire ($P_H = 1$). Since all groups produce this good (guaranteed by a sufficiently large), nominal wages equalize: $w_s = 1$ for all s. Real wages $w_s / P_{D,s}^{1-a}$, however, differ across groups because identity barriers generate group-specific price indices—groups with smaller markets face fewer varieties, higher prices, and lower real income (Section 7).

Differentiated good sector. A continuum of firms in each group use labour as the only input with CRS technology $y = z\ell$.

Key departure: A firm from group s selling in market s' produces using group-s' labour. This assumption reflects a growing body of evidence that firms' hiring tracks the demographics of their customer base. Holzer and Ihlanfeldt (1998) show that the racial composition of a firm's customers predicts the racial composition of its workforce. Kline et al. (2022) document systematic employer-level discrimination in hiring at large U.S. firms, and Rubinstein (2025) shows that customer discrimination directly shapes firms' hiring and employment decisions, particularly in contact-intensive sectors. In the Indian context, Fact 3 confirms this channel: firms' caste-specific employee shares track the caste-specific consumption shares in their local market. The unit cost is:

$$\mathcal{C}_{s,s'} = \begin{cases} w_{s'} & \text{if } s = s', \\ (1 + g_{k,s,s'})\, w_{s'} & \text{if } s \neq s', \end{cases} \tag{10}$$

where $g_{k,s,s'} \geq 0$ is a pair-specific hiring-cost wedge capturing the additional cost of employing out-group workers. A growing body of evidence documents such costs: ethnic diversity on the production floor reduces productivity in Kenyan firms (Hjort, 2014), caste and religious identity create frictions among co-workers in Indian factories (Afridi et al., 2024; Ghosh, 2024), and cross-group hiring is costlier when managers are biased (Glover et al., 2017; Giuliano et al., 2009).[10] Serving market s' requires a per-period fixed cost paid in group-s' labour:

$$f_{s,s'} = \begin{cases} f_d \cdot w_{s'} & \text{if } s = s' \quad \text{(domestic)}, \\ c_{x,s,s'} \cdot w_{s'} & \text{if } s \neq s' \quad \text{(export)}. \end{cases} \tag{11}$$

Entry requires a sunk cost $f_e \cdot w_s$ paid in own-group labour. CES markup pricing gives $p(z, s, s') = \frac{\sigma}{\sigma-1} \cdot \frac{\mathcal{C}_{s,s'}}{z}$.

[10]See Lang and Lehmann (2012) for a survey of racial discrimination in hiring. Cross-group frictions may also arise from network-based hiring that favours own-group candidates or premia demanded by out-group workers.

5.1.3 Firm Revenue and Market Access

A firm from group s with productivity z selling in market s' earns revenue:

$$r(z, s, s') = \Psi_{s',s} \left(\frac{\sigma}{\sigma - 1} \cdot \frac{\mathcal{C}_{s,s'}}{z \cdot P_{D,s'}} \right)^{1-\sigma} (1 - a)(I_{s'} - \bar{h}). \tag{12}$$

Since s now denotes the producer and s' the consumer market, we write $\Psi_{s',s}$ to maintain the convention that the first subscript is the consumer group and the second is the producer group. Revenue depends on income above subsistence, $(I_{s'} - \bar{h})$, not total income. The firm serves market s' if and only if variable profit covers the fixed cost:

$$\frac{r(z^*_{s,s'}, s, s')}{\sigma} = f_{s,s'}. \tag{13}$$

5.2 Equilibrium

Firms draw productivity z from a bounded Pareto distribution with CDF $G(z) = [1 - (z_{\min}/z)^\eta]/[1 - (z_{\min}/z_{\max})^\eta]$ on $[z_{\min}, z_{\max}]$, where $\eta > 0$ is the shape parameter. The finite upper bound ensures all moments are finite without requiring the standard restriction $\eta > \sigma - 1$.

5.2.1 Export Cutoff

We refer to cross-caste sales as "exports" by analogy with international trade, since the model treats caste groups as segmented markets separated by bilateral frictions. Dividing the export zero-profit condition by the domestic one for a group-s firm selling in market s' yields the bilateral export cutoff (see Appendix B for derivation):

$$z^*_{x,s,s'} = z^*_{d,s} \cdot (1 + g_{k,s,s'}) \cdot \left(\frac{c_{x,s,s'}}{f_d} \right)^{\frac{1}{\sigma-1}} \cdot e^{\frac{\beta_{s',s}}{\sigma-1}} \cdot \left(\frac{\kappa_s}{\kappa_{s'}} \right)^{\frac{1}{\sigma-1}}, \tag{14}$$

where κ_s is the market demand index defined in (9). The export cutoff exceeds the domestic cutoff because three frictions—the hiring wedge $(1 + g_{k,s,s'}) > 1$, higher fixed costs $c_{x,s,s'} > f_d$, and demand bias $e^{\beta_{s',s}/(\sigma-1)} > 1$—raise the productivity threshold for cross-group selling. The term $(\kappa_s/\kappa_{s'})^{1/(\sigma-1)}$ captures relative market size: when the destination market s' is larger ($\kappa_{s'} > \kappa_s$), the cutoff falls, making entry easier. Since all parameters are pair-specific, the export cutoff varies across bilateral routes. Only the most productive firms serve out-group markets.

The bilateral frictions can be collapsed into a single *caste-resistance* term. Define:

$$T_{s,s'} \equiv e^{\beta_{s',s}/(\sigma-1)} \cdot (1 + g_{k,s,s'}), \qquad T_{s,s} = 1. \tag{15}$$

This iceberg-equivalent cost converts the demand-side taste bias into a commensurable cost, so that $T_{s,s'}^{1-\sigma} = e^{-\beta_{s',s}} \cdot (1 + g_{k,s,s'})^{1-\sigma}$. The export cutoff (14) becomes:

$$z^*_{x,s,s'} = z^*_{d,s} \cdot T_{s,s'} \cdot \left(\frac{c_{x,s,s'}}{f_d}\right)^{\frac{1}{\sigma-1}} \cdot \left(\frac{\kappa_s}{\kappa_{s'}}\right)^{\frac{1}{\sigma-1}}. \tag{16}$$

5.2.2 Price Index

Let M_s denote the mass of incumbent (active) producers in group s. In a stationary equilibrium with exit rate δ, the mass of entrants satisfies $M_s^e = \delta\, M_s / [1 - G(z^*_{d,s})]$. The CES price index in market s aggregates contributions from domestic firms and exporters from the other two groups:

$$P_{D,s}^{1-\sigma} = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left[M_s\, \Phi_{s,s} + \sum_{s' \neq s} M_{s'}\, e^{-\beta_{s,s'}}\, (1 + g_{k,s',s})^{1-\sigma}\, \Phi_{s',s} \right], \tag{17}$$

where $\Phi_{s',s} \equiv \frac{1}{1-G(z^*_{d,s'})} \int_{z^*_{s',s}}^{z_{\max}} z^{\sigma-1}\, dG(z)$, with $z^*_{s,s} = z^*_{d,s}$ for domestic firms and $z^*_{s',s} = z^*_{x,s',s}$ for exporters ($s' \neq s$).

5.2.3 Free Entry

In a stationary equilibrium with exit rate δ, the free-entry condition for group s equates expected profits to entry costs:

$$\underbrace{\bar{\pi}_{s,s} \cdot [1 - G(z^*_{d,s})]}_{\text{expected domestic profit}} + \sum_{s' \neq s} \underbrace{\bar{\pi}_{s,s'} \cdot [1 - G(z^*_{x,s,s'})]}_{\text{expected profit in market } s'} = \delta\, f_e\, w_s, \tag{18}$$

where $\bar{\pi}_{s,s'}$ is the average operating profit of a group-s firm conditional on serving market s'. Given the cutoff ratios (14), this determines the domestic cutoff $z^*_{d,s}$ for each group (see Appendix B for the expression under Pareto). In a stationary equilibrium, total operating profits of group-s firms exactly equal the flow of entry costs $M_s^e f_e w_s$—where M_s^e is the mass of group-s entrants per period—which are paid in own-group labour. Net profits are therefore zero ($\Pi_s = 0$), and group income reduces to $I_s = w_s L_s = L_s$.

5.2.4 Market Clearing

D-sector goods market. In each market s, expenditure on differentiated goods equals total revenue earned by all firms selling in s:

$$(1 - a)(I_s - \bar{h}) = M_s\, \bar{R}_{s,s} + \sum_{s' \neq s} M_{s'}\, \bar{R}_{s',s}, \tag{19}$$

where $\bar{R}_{s',s}$ is the average revenue earned in market s per active group-s' firm (averaging over all group-s' incumbents, including those that do not serve market s, for whom $\bar{R}_{s',s} = 0$). This pins down the mass of active firms M_s.

Labour market. Group-s labour is demanded by: (i) homogeneous-good production, (ii) all differentiated-good production destined for market s (from firms of all groups, since production uses destination labour), (iii) fixed costs of serving market s, and (iv) entry costs for group-s firms:

$$L_s = H_s^{\text{prod}} + \sum_{s'} M_{s'} \left[\frac{\sigma - 1}{\sigma} \frac{\bar{R}_{s',s}}{\mathcal{C}_{s',s}} + \frac{1 - G(z^*_{s',s})}{1 - G(z^*_{d,s'})} \frac{f_{s',s}}{w_s} \right] + M_s^e f_e. \tag{20}$$

The summation runs over all groups s' (including $s' = s$). Variable labour demand in market s by group-s' firms equals $(\sigma - 1)/\sigma$ times revenue divided by the unit cost $\mathcal{C}_{s',s}$, which includes the hiring wedge $g_{k,s',s}$ for $s' \neq s$. Fixed-cost labour equals the conditional fraction of active group-s' firms that serve market s—i.e., $[1 - G(z^*_{s',s})]/[1 - G(z^*_{d,s'})]$, which equals one when $s' = s$—times the fixed cost $f_{s',s}/w_s$ in labour units. Entry costs $M_s^e f_e w_s$ are paid in own-group labour.

5.2.5 Real Income

With non-homothetic preferences, the real income measure uses income above subsistence:

$$\text{Real income}_s = \frac{I_s - \bar{h}}{P_{D,s}^{1-a}}, \qquad \text{Real income p.c.}_s = \frac{(I_s - \bar{h})/L_s}{P_{D,s}^{1-a}}. \tag{21}$$

Groups with higher Income per-capita $(I_s - \bar{h})/L_s$ enjoy disproportionately higher real income, since they allocate a larger share of income to differentiated goods where variety gains accrue.

6 Quantitative Analysis

We calibrate the model to match salient features of India's economy. The calibrated model is then used to study how identity-based frictions shape firm behaviour, market structure, and real income across caste groups.

6.1 Calibration

The model is parameterised in two stages. First, a set of parameters are determined externally using data or standard values from the literature. Second, the remaining parameters—governing cross-caste frictions—are jointly calibrated to match targeted moments from the MSME data.

6.1.1 Externally Set Parameters

The elasticity of substitution is set to $\sigma = 5$, a standard value in the trade literature (Broda and Weinstein, 2006).[11] The firm exit rate is $\delta = 0.088$, and the entry cost is normalised to $f_e = 1$. Labour endowments L_s capture cross-caste differences in effective labour supply and are set to $L_{\text{LC}} = 0.680$, $L_{\text{MC}} = 0.830$, and $L_{\text{HC}} = 1.000$, reflecting the MPCE of each caste group relative to the HC group in Table A.4. Since income in the model equals $I_s = L_s$ (see below), matching L_s to relative MPCE ensures that the model reproduces the observed cross-caste income ranking and the relative size of each group's market.

Share of homogeneous-good sector a. The homogeneous-good expenditure parameter is $a = 0.30$ and the subsistence requirement is $\bar{h} = 0.30$. There is no direct measure of a—the share of expenditure on products where identity barriers do not apply—in the literature. We estimate it by repeating the regressions presented in Section 4.3 for each product separately and computing the share of expenditure where the demand–employment coefficients are positive and significant.

Given a, we set $\bar{h} = 0.30$, so that total expenditure on the homogeneous-good sector is 51% for the HC group (larger for LC and MC groups). The combination of $\{a, \bar{h}\}$ implies a lower bound on the cost of identity barriers for two reasons. First, a high $\{a, \bar{h}\}$ means a lower share of the distorted sector in the economy. Second, a high subsistence requirement $\bar{h}$ raises the relative revenue elasticity of LC firms for a given demand shock, so the calibration strategy implies a smaller demand bias β. We use different levels of $\{a, \bar{h}\}$ and provide bounds on income losses in Section C.2 in the appendix.

6.1.2 Internally Calibrated Parameters

The remaining 15 parameters are internally calibrated by matching targeted moments in the data to their counterparts in the model. We group them into four sets, each with a distinct identification source.

Firm-size distribution: η and f_d. Two parameters govern the firm-size distribution. A decline in the Pareto shape parameter η makes the tail of the productivity distribution thicker, increasing the revenue share of the top 10% of firms; we therefore

[11] We use a bounded Pareto distribution to ensure all moments remain finite despite $\eta < \sigma - 1$, as empirical estimates of $\eta/(\sigma - 1)$ are below one in many sectors. The results are robust to alternative values: recalibrating all 13 parameters at $\sigma = 3$ and $\sigma = 4$ yields aggregate gains from removing all identity frictions of 1.116 and 1.095, respectively, compared with 1.076 at $\sigma = 5$. Lower σ raises the cost of segmentation because varieties are less substitutable, but the qualitative pattern—hiring wedges dominating taste bias—is unchanged.

match η to this moment (data: 87.7%, model: 87.75%). Conversely, an increase in the domestic fixed cost f_d raises the share of output produced by the bottom half of the revenue distribution, as it compresses the lower tail of the firm-size distribution; we match f_d to the revenue share of the bottom 50% of firms (data: 1.4%, model: 1.40%).

Taste elasticity β. The main parameter of interest is the demand-bias parameter β, which governs the discount consumers apply to goods from other castes ($\Psi = e^{-\beta}$). We calibrate β using the reduced-form revenue elasticity estimates from Section 4.4. The identification relies on the following partial-equilibrium decomposition. Holding firm selection and price indices fixed, the average revenue growth of group-s firms decomposes exactly as:

$$\frac{\bar{R}_s^{PE}}{\bar{R}_s} = \sum_{s'} \bar{\omega}_{s,s'} \cdot \frac{E_{D,s'}^{PE}}{E_{D,s'}}, \tag{22}$$

where $\bar{\omega}_{s,s'} \equiv \bar{R}_{s,s'} / \bar{R}_s$ is the share of group-s firms' revenue earned from market s' and $E_{D,s'} = (1-a)(I_{s'} - \bar{h})$ is group-s' differentiated-good expenditure. Revenue growth is a revenue-share-weighted average of differentiated-good expenditure growth across destination markets (see Appendix C.1 for derivation).

The demand-bias parameter β governs the revenue shares $\bar{\omega}_{s,s'}$: higher β raises own-market concentration ($\bar{\omega}_{s,s}$ increases in β), since the demand discount $e^{-\beta}$ compresses cross-group revenue. Consequently, when LC demand rises, LC firms—which earn a larger share of revenue from the LC market—benefit more than HC firms. This differential response grows monotonically in β (Figure 5). We feed the empirically estimated demand shocks (a 9.2% increase in LC income and a 4.5% increase in MC income, from the rainfall IV in Table 3) and require that the model-implied relative revenue growth of LC-owned firms matches the empirical estimate from Figure 4 (estimates in red).

A natural concern is that the rainfall shock is temporary whereas the model is static. The partial-equilibrium decomposition in Equation (22) addresses this: it decomposes revenue growth into a weighted sum of expenditure growth across destination markets, where the weights $\bar{\omega}_{s,s'}$ are steady-state revenue shares determined by the structural parameters β, g_k, and c_x. Because these weights are properties of the stationary equilibrium—not of the shock's duration—the mapping from the empirical revenue elasticity to β does not depend on whether the demand shock is temporary or permanent. What matters is the cross-sectional difference in revenue exposure to the LC market between LC-owned and HC-owned firms, which reflects the steady-state degree of demand segmentation.

The calibrated value is $\beta = 0.284$, implying $\Psi = e^{-0.284} \approx 0.753$. The symmetry restriction $\beta_{s,s'} = \beta$ for all $s \neq s'$ is imposed because the rainfall IV identifies a single differential revenue response (LC relative to HC), which pins down one taste-bias parameter; asymmetric $\beta_{s,s'}$ across all six bilateral pairs would require additional identifying moments—for instance, separate revenue elasticities for each origin-destination pair—that the cross-sectional design does not provide. The restriction is conservative for the demand-side channel: if the caste hierarchy implies $\beta_{\text{HC,LC}} > \beta_{\text{LC,HC}}$, as one would expect when higher-ranked groups discriminate more against lower-ranked goods, an asymmetric calibration would attribute a larger share of the total barrier to demand-side bias and a correspondingly smaller share to hiring wedges. Differential market access across bilateral pairs is instead generated by the 12 supply-side parameters (g_k and c_x), which are fully asymmetric.

Hiring-cost wedges $g_{k,s,s'}$. The 6 off-diagonal elements of the hiring-cost matrix are calibrated to match the worker-composition shares across castes in Table 2. A higher wedge $g_{k,s,s'}$ raises the marginal cost of hiring group-s' workers for group-s firms, reducing the employment share of group-s' workers in those firms. For example, the share of MC workers employed by LC-owned firms (20.27% in the data) pins down $g_{k,\text{LC}\to\text{MC}}$: a higher wedge would lower this share by making MC labour more expensive for LC firms.

Export fixed costs $c_{x,s,s'}$. The 6 off-diagonal elements of the export fixed-cost matrix are calibrated to match the hiring participation rates—the fraction of firms of each caste that employ at least one worker from each other caste (Table 2, Extensive Margin). In the model, a firm from group s hires group-s' workers if and only if it finds it profitable to sell in market s', which requires its productivity to exceed the cutoff $z^*_{s,s'}$. A higher fixed cost $c_{x,s,s'}$ raises this cutoff, reducing the fraction of group-s firms that hire group-s' workers. For example, only 13.61% of MC-owned firms hire LC workers; this low participation rate implies a relatively high $c_{x,\text{MC}\to\text{LC}}$. Table 5 reports all parameter values and the targeted moments with their data counterparts.

6.1.3 Discussion on Calibrated Parameters

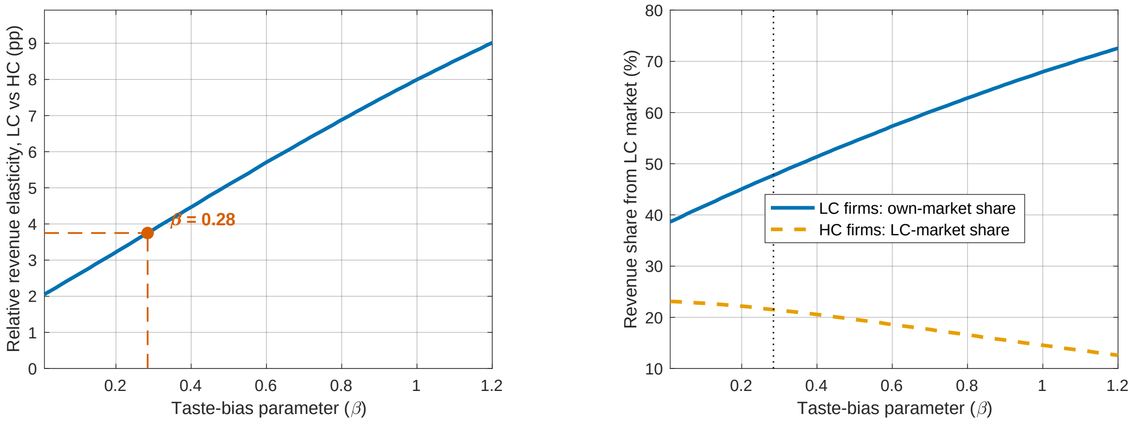
Several features of the calibrated parameters merit discussion. First, the hiring-cost wedges g_k range from 0.047 to 0.177. MC firms face the highest wedge when hiring LC workers ($g_{k,\text{MC}\to\text{LC}} = 0.177$), while LC firms face the lowest wedge when accessing HC markets ($g_{k,\text{LC}\to\text{HC}} = 0.047$). These wedges capture the combined effect of network frictions, information asymmetries, and potential discrimination in cross-caste hiring.

Second, the export fixed costs $c_{x,s,s'}$ range from 0.130 to 0.143, implying ratios to the domestic fixed cost ($f_d = 0.020$) that are consistent with the trade literature's estimates of export-to-domestic fixed-cost ratios. HC firms face the lowest export fixed costs ($c_{x,\text{HC}\to\text{MC}} = 0.045$ and $c_{x,\text{HC}\to\text{LC}} = 0.087$), suggesting that high-caste firms encounter fewer barriers to cross-caste market access.

Table 5: Calibrated Parameters and Targeted Moments

Panel A: Externally Set Parameters					
Parameter	Description	Value	Source/Target		
σ	Elast. of substitution	5.000	Trade literature		
a	H-good expend. share	0.300	Expenditure data		
$\bar{h}$	Subsistence req.	0.300	Engel curve estimation		
δ	Firm exit rate	0.088	Literature		
f_e	Entry cost	1.000	Normalisation		
L_{LC}	LC labour endowment	0.680	Relative income, NSS		
L_{MC}	MC labour endowment	0.830	Relative income, NSS		
L_{HC}	HC labour endowment	1.000	Normalisation		
Panel B: Internally Calibrated Parameters					
Parameter	Description	Value	Targeted Moment	Data	Model
β	Demand bias	0.284	Rel. revenue elasticity (LC vs HC)	3.60	3.75
Firm-size distribution parameters					
η	Pareto shape	3.100	Revenue share, top 10%	87.7	87.75
f_d	Domestic fixed cost	0.020	Revenue share, bottom 50%	1.4	1.40
Hiring-cost wedge $g_{k,s,s'}$: matched to worker-composition shares (%)					
$g_{k,\text{LC}\to\text{MC}}$	LC $\to$ MC	0.167	MC workers in LC firms	20.27	20.20
$g_{k,\text{LC}\to\text{HC}}$	LC $\to$ HC	0.047	HC workers in LC firms	30.06	30.18
$g_{k,\text{MC}\to\text{LC}}$	MC $\to$ LC	0.177	LC workers in MC firms	15.35	15.47
$g_{k,\text{MC}\to\text{HC}}$	MC $\to$ HC	0.135	HC workers in MC firms	20.43	20.49
$g_{k,\text{HC}\to\text{LC}}$	HC $\to$ LC	0.121	LC workers in HC firms	20.52	20.46
$g_{k,\text{HC}\to\text{MC}}$	HC $\to$ MC	0.139	MC workers in HC firms	24.64	24.64
Export fixed cost $c_{x,s,s'}$: matched to hiring participation rates (%)					
$c_{x,\text{LC}\to\text{MC}}$	LC $\to$ MC	0.156	% LC firms hiring MC	18.66	18.23
$c_{x,\text{LC}\to\text{HC}}$	LC $\to$ HC	0.130	% LC firms hiring HC	25.63	24.77
$c_{x,\text{MC}\to\text{LC}}$	MC $\to$ LC	0.143	% MC firms hiring LC	13.61	13.97
$c_{x,\text{MC}\to\text{HC}}$	MC $\to$ HC	0.089	% MC firms hiring HC	21.93	21.48
$c_{x,\text{HC}\to\text{LC}}$	HC $\to$ LC	0.087	% HC firms hiring LC	23.51	23.09
$c_{x,\text{HC}\to\text{MC}}$	HC $\to$ MC	0.045	% HC firms hiring MC	38.71	39.76

Notes. Panel A reports parameters set externally from data or the literature. Panel B reports parameters internally calibrated by matching targeted moments from the MSME data. The "Data" column reports the empirical moment; the "Model" column reports the model-implied moment at the calibrated parameter values. Revenue shares are from the MSME firm-size distribution. Worker-composition shares are from the MSME cross-section. Hiring participation rates are the fraction of firms employing at least one worker of the indicated caste (Table 2, Extensive Margin). The relative revenue elasticity is the partial-equilibrium revenue response of LC-owned firms relative to HC-owned firms following the estimated caste-specific demand shocks (Table A.12).



(a) Relative revenue elasticity

(b) Revenue share from LC market

Figure 5: Identification of β: Taste Bias and Revenue Elasticity

Notes. Panel (a) plots the partial-equilibrium relative revenue elasticity (LC vs HC firms) against the taste-bias parameter β, holding all other parameters at their calibrated values. At each β, we solve the model equilibrium and compute the revenue response to caste-specific demand shocks (9.2% for LC, 4.5% for MC). The red marker denotes the calibrated $\beta = 0.284$. Panel (b) shows the revenue share that LC and HC firms earn from the LC market as a function of β. Higher β increases LC firms' own-market concentration while reducing HC firms' LC-market penetration.

Third, the demand-bias parameter $\beta = 0.284$ implies that cross-caste demand is discounted by a factor of $e^{-0.284} \approx 0.753$ relative to own-caste demand. This bias, combined with the production-side frictions, generates the observed patterns of market segmentation along caste lines.

6.1.4 Overidentification: Firm Size and Workforce Diversity

As an external check on the calibrated parameters, we compare the model's predictions for an untargeted moment: the gradient of the own-caste employee share with respect to firm size. This gradient is not used in calibration—the targeted moments are worker-composition *levels*, hiring participation *rates*, and the revenue *elasticity*—yet it emerges endogenously from the interaction of β and g_k. In the model, firms grow by serving cross-caste markets, which under destination labour requires hiring workers from those markets, mechanically reducing the own-caste share. The steepness of this decline reflects how quickly the demand penalty and hiring costs are overcome as productivity rises.

In the data, regressing the own-caste employee share (%) on log(employees) with caste, district, and NIC-4 fixed effects yields a pooled slope of -10.2 pp per log-unit of employment (Online Appendix Table A.8, Column 1). In the model, we simulate a cross-section of firms and run the analogous OLS regression; the pooled slope is -10.6 pp. The close match is reassuring, since this moment jointly disciplines both the demand bias (which determines how much cross-caste revenue firms can earn) and the hiring wedges (which determine the cost of the workforce diversification needed to access those markets).

A second untargeted check concerns the distribution of aggregate revenue across caste groups. In the data, HC-owned firms account for 76.9% of total revenue, MC-owned firms for 15.5%, and LC-owned firms for 7.6% (Table 1). The model generates shares of 68.4%, 20.1%, and 11.5%, respectively. The model correctly captures HC dominance—driven primarily by the larger mass of HC firms—though it understates the HC share by about 8 pp.

6.2 Ad Valorem Interpretation of Cross-Caste Frictions

The model features two types of bilateral frictions: a demand-side identity barrier (β) and a supply-side hiring-cost wedge ($g_{k,s,s'}$). These operate through different channels—the former shifts demand, the latter raises marginal cost—but both reduce a firm's revenue from cross-caste markets. A natural question is: how large are these frictions, expressed in common units? In this section, we convert both frictions into ad valorem equivalents, analogous to the iceberg transportation costs estimated in the trade literature (see Section C.3).

The demand-side identity barrier $\Psi_{s',s} = e^{-\beta}$ enters revenue multiplicatively. In a CES demand system, an iceberg cost τ reduces revenue by $\tau^{1-\sigma}$. Equating this to the demand discount, $\tau^{1-\sigma} = e^{-\beta}$, gives the ad valorem equivalent $\tau - 1 = e^{\beta/(\sigma-1)} - 1$. At the calibrated values $\beta = 0.284$ and $\sigma = 5.000$, the ad valorem equivalent is 7.4%—that is, the demand-side identity barrier has the same revenue effect as a 7.4% iceberg transportation cost.

The hiring-cost wedge $g_{k,s,s'}$ is already expressed in iceberg form. When a group-s firm hires workers from group s', it pays $(1 + g_{k,s,s'}) \cdot w_{s'}$ per unit of labour, but only $w_{s'}$ worth of productive labour services is delivered. The fraction $g_{k,s,s'}$ is a deadweight loss—real resources consumed by employer–employee preferences, network frictions, and information asymmetries in cross-caste hiring—so the ad valorem equivalent is simply $g_{k,s,s'}$ itself. Crucially, the wedge is *asymmetric* across bilateral routes: the highest is $g_{k,\mathrm{MC}\to\mathrm{LC}} = 0.177$ (17.7%), the lowest is $g_{k,\mathrm{LC}\to\mathrm{HC}} = 0.047$ (4.7%), and the average across all six routes is 13.1%.

The two frictions combine multiplicatively into the composite caste resistance defined in equation (15):

$$T_{s,s'} = \underbrace{e^{\beta/(\sigma-1)}}_{\substack{\text{demand bias} \\ (7.4\%)}} \times \underbrace{(1 + g_{k,s,s'})}_{\substack{\text{hiring wedge} \\ (4.7\text{–}17.7\%)}}. \tag{23}$$

The multiplicative structure means the two frictions compound. Table 6 reports the full decomposition for each bilateral route. Two patterns stand out. First, the hiring wedge dominates: the demand-bias component contributes a uniform 7.4% across all routes, while the hiring wedge ranges from 4.7% to 17.7%. Second, the composite barriers are comparable to interstate transportation costs in India—ranging from 12.4% to 26.4% ad valorem, within the range estimated by Asturias et al. (2019) for Indian manufacturing for internal trade barriers across Indian states.

Table 6: Ad Valorem Decomposition of Caste Resistance $T_{s,s'}$

Route ($s \to s'$)	β	$g_{k,s,s'}$	β component $e^{\beta/(\sigma-1)}$	g_k component $(1+g_k)$	Composite $T_{s,s'}$	Ad valorem (%)
LC $\to$ MC	0.284	0.167	1.074	1.167	1.253	25.3
LC $\to$ HC	0.284	0.047	1.074	1.047	1.124	12.4
MC $\to$ LC	0.284	0.177	1.074	1.177	1.264	26.4
MC $\to$ HC	0.284	0.135	1.074	1.135	1.219	21.9
HC $\to$ LC	0.284	0.121	1.074	1.121	1.204	20.4
HC $\to$ MC	0.284	0.139	1.074	1.139	1.223	22.3
Mean		0.131			1.215	21.5

Notes. The demand-bias component $e^{\beta/(\sigma-1)} = e^{0.284/4} \approx 1.074$ is common to all pairs since β is calibrated symmetrically. The hiring-wedge component $(1 + g_{k,s,s'})$ varies by bilateral pair. The composite $T_{s,s'} = e^{\beta/(\sigma-1)} \times (1 + g_{k,s,s'})$, with $\tau_{s,s'} = 1$ in the baseline (no iceberg transportation costs). Ad valorem $= T_{s,s'} - 1$. Parameter values from Table 5.

6.3 Results

6.3.1 Firm Size, Market Access and Out-Caste Hiring

We use the calibrated model to examine how the taste-bias parameter β shapes firm size, cross-caste market access, and hiring composition. At the calibrated $\beta = 0.284$, cross-caste demand is discounted by a factor $\Psi = e^{-\beta} \approx 0.753$ relative to own-caste demand.

Figure 6(a) plots three aggregate outcomes—average firm revenue (a proxy for firm size), the share of firms selling to at least one other caste, and the share of out-group employees—as a function of β, with each series normalised to one at the calibrated value $\beta = 0.284$. All three outcomes decline monotonically in β. This stems from two reinforcing channels. First, a higher β reduces the demand that consumers direct toward goods produced by other castes, shrinking the revenue that firms can earn from cross-caste markets. Only the most productive firms find it profitable to bear the fixed cost $c_{x,s,s'}$ and sell to a socially distant group, so the share of firms participating in cross-caste trade falls. Second, because firms in the destination-labour framework hire workers from the market they sell into, fewer cross-caste sales translate directly into fewer out-group employees. As β rises, the share of firms trading out-caste and the out-group employment share decline sharply, while average firm size falls more gradually (Figure 6(a)).

The asymmetry between the firm-size response and the trading/hiring responses reflects the model's selection mechanism. A marginal increase in β has a small effect on the revenue of infra-marginal firms that already concentrate on their home market, but a large effect on the extensive margin—pushing firms near the export cutoff $z^*_{x,s,s'}$ out of cross-caste markets. This selection effect is the primary driver of the decline in out-caste hiring participation.

Free entry, firm size, and reallocation. At first glance, it may seem puzzling that average firm size changes with β when the free-entry condition holds. In our multi-market setting, the free-entry condition links average *total* revenue to the average *total* fixed cost paid by a surviving firm:

$$\bar{r}_{\text{total}} = \sigma\left(\bar{f}_{\text{total}} + \frac{\delta f_e}{\text{succ}}\right), \tag{24}$$

where $\text{succ} = 1 - G(z^*_{d,s})$ is the probability of drawing a productivity above the domestic cutoff, and $\bar{f}_{\text{total}} = f_d + \sum_{s' \neq s} \Pr(z \geq z^*_{x,s,s'} \mid z \geq z^*_{d,s}) \cdot c_{x,s,s'}$ is endogenous. When β rises, the export cutoffs $z^*_{x,s,s'}$ increase and fewer firms pay the cross-caste fixed costs, so $\bar{f}_{\text{total}}$ falls and free entry requires lower average total revenue—i.e., smaller firms on average.

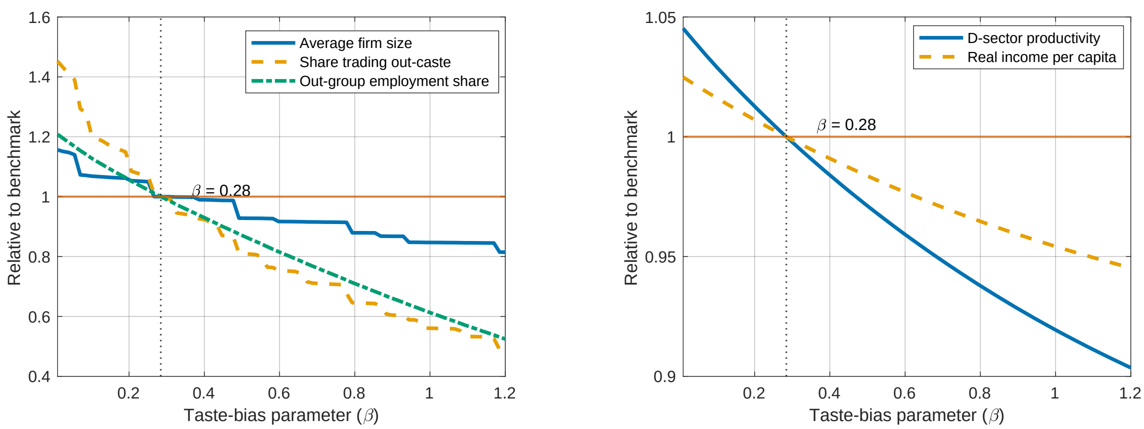
Decomposing average revenue into domestic and export components clarifies the mechanism. Selection on the export margin is *stronger* at higher β: the firms that *do* export are more productive and individually larger, and average domestic revenue per firm also rises because fewer imported varieties reduce competitive pressure in the home market. But the *composition* of the firm population shifts toward domestic-only firms, which earn revenue from a single market. This composition effect dominates, and average total revenue falls.

This composition shift constitutes a form of *misallocation through market fragmentation*. In the frictionless benchmark ($\beta = 0$, $g_k = 0$), the most productive firms serve all three markets and grow large, while the least productive are selected out. As β rises, productive firms that could profitably serve cross-caste markets are pushed below the export cutoff, shrinking their scale and releasing labour. These inputs are absorbed by the entry of additional domestic-only firms—firms that are individually less productive and serve a smaller market. Labour is thus *reallocated from large, multi-market firms toward small, single-market firms*, compressing the upper tail of the firm-size distribution. This is precisely the configuration that Restuccia and Rogerson (2008) show to be most damaging for aggregate productivity: distortions positively correlated with firm size, penalising the most productive firms. Identity-based segmentation is also a concrete example of the size-dependent distortions that Hsieh and Klenow (2014) identify as hindering the growth of Indian firms.

6.3.2 Aggregate Productivity and Income per Capita

We next examine how taste bias affects aggregate productivity and real income. Figure 6(b) plots two measures against β, each normalised to one at the calibrated value: (i) D-sector productivity, defined as real differentiated-good output per unit of D-sector labour; and (ii) real income per capita, measured using the non-homothetic cost-of-living index $P_{D,s}^{1-a}$.

Both measures decline monotonically as β rises, but they differ in magnitude (Figure 6(b)). Higher β reduces cross-caste market access, lowering the mass of firms selling in each market. This raises the D-sector price index $P_{D,s}$ through two channels: fewer imported varieties (the variety effect) and weaker competitive selection among the surviving firms (the selection effect). Real income per capita falls by less than D-sector productivity because the non-homothetic deflator $P_{D,s}^{1-a}$ discounts the D-sector price index by the differentiated-good expenditure share $(1-a)$.



(a) Firm size, trading and hiring

(b) Productivity and real income

Figure 6: Taste Bias, Firm Size, Market Access and Aggregate Productivity

Notes. Panel (a) plots three normalised outcomes against β: average firm revenue (blue solid), the share of firms selling to at least one other caste (orange dashed), and the aggregate out-group employment share (teal dash-dotted). Panel (b) plots D-sector productivity $\sum_s Q_{D,s} / \sum_s L_{D,s}$ (blue solid) and real income per capita $\sum_s (I_s - \bar{h}) / P_{D,s}^{1-a} / \sum_s L_s$ (orange dashed). All series are normalised to one at $\beta = 0.284$.

6.3.3 Identity Barriers, Income and Inequality

Table 7 reports real income per capita for each caste group and the economy-wide aggregate under several counterfactual scenarios, and Figure 7 plots the corresponding efficiency–inequality tradeoffs: movement toward the southeast represents an unambiguous improvement (higher aggregate income, lower cross-caste inequality). Four main findings emerge.

6.3.3.1 Supply-side frictions dominate demand-side bias. Under destination labour, both frictions are bilateral: hiring wedges raise the unit cost of serving a cross-caste market, just as taste bias discounts demand from that market. At the calibrated values, the cost channel dominates: eliminating hiring wedges ($g_k = 0$) raises aggregate real income by 4.7%—roughly twice the 2.6% gain from removing taste bias ($\beta = 0$). In both cases LC gains the most, since its small home market makes LC firms the most reliant on cross-caste channels. Removing both frictions simultaneously yields a 7.6% gain, and inequality falls by 46%. These results suggest that labour-market interventions—strengthening anti-discrimination enforcement, standardised worker certification, cross-caste apprenticeships—are likely to deliver larger real income gains than policies aimed directly at consumer preferences, and that targeting both channels simultaneously is most effective.

6.3.3.2 Market access is asymmetrically valuable. Because LC has the smallest home market, it creates the fewest firms and the least product variety in autarky, making cross-caste exchange most valuable for LC. Full autarky reduces aggregate income by 9.7% (LC -11.4%, HC -10.3%, MC -7.7%). Isolating LC alone costs 5.4% in aggregate, with nearly the entire burden falling on LC. The value of specific bilateral links reflects market size: blocking LC–HC trade costs 2.5%, whereas blocking LC–MC trade costs only 0.1%.

6.3.3.3 General-equilibrium spillovers: barriers between two groups can benefit a third. When LC–MC trade is blocked, higher price indices in those markets raise the revenue per variety available to HC firms, whose access is unaffected. Through free entry, higher expected profits attract additional HC firms, lowering $P_{D,\mathrm{HC}}$ and raising HC real income by 3.1% even though no HC trade channel is directly affected. The effects of segmentation are therefore inherently general-equilibrium and cannot be assessed market by market.

6.3.3.4 Directional counterfactuals as extreme asymmetric taste bias. Although the baseline calibration imposes a symmetric β, the directional counterfactuals in Panel B can be reinterpreted as extreme asymmetric preferences. Blocking trade on a route is equivalent to $\beta_{s,s'} \to \infty$; the "LC buys but cannot sell to HC" scenario therefore corresponds to $\beta_{\mathrm{HC,LC}} = \infty$ while $\beta_{\mathrm{LC,HC}}$ remains at its calibrated value—the hierarchical pattern one would expect if higher-ranked groups discriminate more strongly. That this extreme asymmetry is primarily regressive (20% rise in inequality) rather than costly in aggregate (near-zero income change) is a striking finding. More broadly, the total bilateral barrier combining taste bias and hiring wedges is pinned by the data moments. Removing all identity frictions yields a 7.6% aggregate gain regardless of how the barrier is decomposed between demand and supply sides; what would change under asymmetric β is only the decomposition, not the total cost of segmentation or the distributional burden on LC.

The homogeneous economy, which pools all groups into a single representative caste, provides an upper bound: aggregate real income rises by 57.8 per cent, reflecting the combined benefits of removing all identity-based frictions and creating a single larger market that supports greater firm entry and product variety.

Table 7: Counterfactual Real Income Per Capita

Scenario	LC	MC	HC	Aggregate	Var(log RI)
Panel A: Removing frictions					
Benchmark (calibrated)	1.000	1.000	1.000	1.000	0.0168
No taste bias ($\beta = 0$)	1.038	1.012	1.030	1.026	0.0165
No hiring wedges ($g_k = 0$)	1.067	1.055	1.031	1.047	0.0134
No bias + no wedges ($\beta = 0, g_k = 0$)	1.121	1.109	1.031	1.076	0.0090
Homogeneous economy	1.900	1.634	1.380	1.578	0.0000
Panel B: Blocking inter-caste trade					
Full autarky	0.886	0.923	0.897	0.903	0.0178
LC isolated	0.886	0.978	0.954	0.946	0.0250
LC–HC trade blocked	0.917	1.091	0.923	0.975	0.0210
LC–MC trade blocked	0.985	0.963	1.031	0.999	0.0233
LC buys HC, cannot sell	0.977	1.004	1.008	1.000	0.0201

Notes. The first four columns report real income per capita $(I_s - \bar{h})/P_{D,s}^{1-a}$ for the indicated group, normalised to the benchmark calibration. The last column reports the population-weighted variance of log real income per capita across caste groups: $\text{Var} = \sum_s \omega_s(\log \text{RI}_s - \bar{\mu})^2$, where $\omega_s = L_s / \sum_s L_s$ and $\bar{\mu} = \sum_s \omega_s \log \text{RI}_s$. Panel A removes demand-side taste bias or supply-side hiring wedges; the homogeneous economy pools all groups into a single representative caste. Panel B introduces prohibitively large export fixed costs on selected bilateral routes, shutting down the corresponding inter-caste trade flows. "LC isolated" blocks all trade involving LC firms and consumers. "LC buys HC, cannot sell" blocks LC firms from selling to HC consumers while allowing HC firms to sell to LC consumers.

7 Model Limitations and Extensions

Our framework makes several simplifying assumptions. We discuss how relaxing them could alter the quantitative predictions.

Geography and spatial sorting. The model is aspatial: all groups coexist in a single location. A natural concern is that the estimated barriers partly reflect spatial segregation of caste groups rather than identity-based frictions per se. Three observations address this concern.

First, the taste-bias parameter β is identified from the revenue elasticity in Equation (4), which exploits *within-district* variation—comparing LC-owned to HC-owned firms in the same district facing the same rainfall shock—so spatial sorting *across* districts does not confound the estimate. Moreover, the LC revenue elasticity is stable

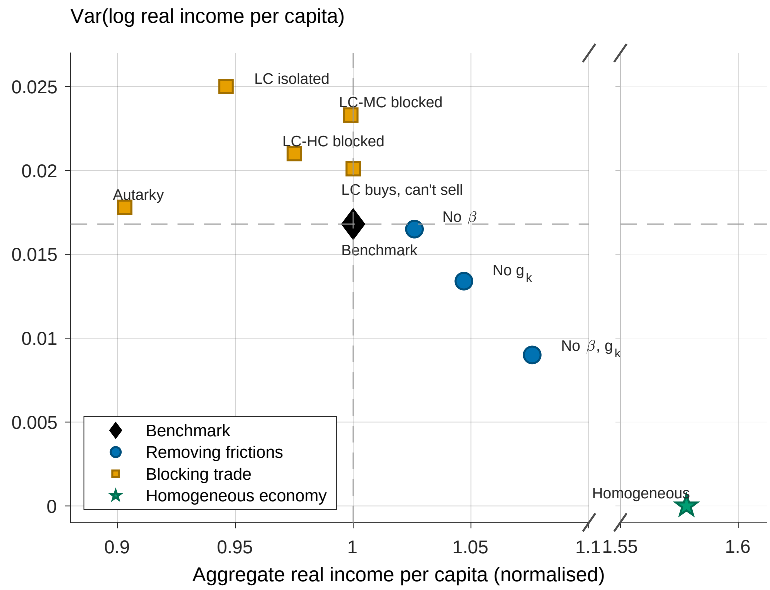


Figure 7: Counterfactual Real Income and Inequality

Notes. Each point represents a counterfactual scenario from Table 7. The horizontal axis plots aggregate real income per capita, normalised to the benchmark calibration. The vertical axis plots the population-weighted variance of log real income per capita across caste groups. Diamonds denote the benchmark, circles denote scenarios that remove identity-based frictions (Panel A), squares denote scenarios that block inter-caste trade (Panel B), and the star denotes the homogeneous economy. Dashed lines mark the benchmark values. Movement toward the southeast (higher income, lower inequality) represents an unambiguous improvement.

when the sample is restricted to residentially integrated districts (Table A.12, column 9), where caste groups share local markets and geographic separation is minimal. The demand channel operates independently of spatial sorting.

Second, we extend the baseline model by introducing iceberg transportation costs τ on all cross-caste trade and recalibrate (Appendix C.3). This provides formal bounds on the role of spatial sorting. We consider two values: a lower bound of $\tau = 1.03$ (3% ad valorem), corresponding to estimated costs between neighbouring Indian states, and a higher bound of $\tau = 1.10$ (10% ad valorem), corresponding to moderately distant states. Appendix C.5.1 provides a detailed discussion of why these values represent reasonable estimates of transportation costs in our setting. Removing all identity frictions ($\beta = 0$, $g_k = 0$) raises aggregate real income to 1.061 times the benchmark under $\tau = 1.03$, and to 1.039 times the benchmark under $\tau = 1.10$—roughly half the aspatial baseline gain of 1.076. Transportation costs therefore attenuate but do not eliminate the real income gains from removing identity-based frictions.

Third, treating spatial segregation as fully exogenous may partial out part of the very barrier we aim to measure. If caste-based preferences over neighbours or discriminatory housing markets sort groups into distinct locations, then the iceberg cost τ is partly a *consequence* of identity frictions—not an independent confound. In that case, the counterfactual that removes β and g_k while holding τ fixed understates the true cost of identity barriers, because eliminating caste-based preferences would also reduce spatial segregation and hence τ. Disentangling the endogenous and exogenous components of spatial segregation—by modelling residential choice alongside product market segmentation—is an important avenue for future research.

Equalised wages. The homogeneous good sector serves as the numeraire and equalises nominal wages across castes ($w_s = 1$). In practice, substantial caste-based wage gaps persist in India, reflecting differences in human capital, occupational sorting, and discrimination (Madheswaran and Attewell, 2007; Hnatkovska et al., 2012). If wages were allowed to differ—for instance, by dropping the homogeneous good sector or by introducing group-specific labour productivity—the gains from removing hiring wedges would likely be larger, as the wedge compounds a pre-existing wage disadvantage for LC workers. At the same time, the real income ranking across counterfactuals would depend on general-equilibrium wage adjustments, which our framework abstracts from.

Capital and financial frictions. The model features labour as the sole factor of production and is therefore silent on capital misallocation. A growing literature documents that caste-based networks distort the allocation of capital (Banerjee and Munshi, 2004; Goraya, 2023), and that lower-caste entrepreneurs face tighter credit constraints. Introducing capital alongside labour would add a complementary channel: identity-based barriers would suppress both labour demand *and* capital accumulation for firms selling across caste lines. Our counterfactual real income gains should therefore be interpreted as a lower bound on the full cost of caste-based segmentation. A related identification concern is that rainfall shocks may relax financial constraints for LC entrepreneurs through intra-caste lending networks, rather than operating through demand. Three pieces of evidence weigh against this channel: the revenue response is larger among big firms (Table A.12, column 12), which are less likely to be credit-constrained; the competition and contact-intensity interactions in Fact 3—which should not vary with financial frictions—remain significant; and the foreign-demand placebo, which is expansionary but caste-neutral, produces no differential hiring of LC workers.

Within-caste heterogeneity. We aggregate thousands of *jatis* into three broad groups (LC, MC, HC). As Boken et al. (2025) and Fujiy et al. (2025) show, trade frictions operate at the jati level as well. Our three-group aggregation therefore understates the extent of market segmentation. A finer partition would likely amplify the real income costs of identity barriers, since the effective number of segmented markets is much larger than three.

Endogenous identity and social change. The taste-bias parameter β is fixed in our model. In reality, preferences may evolve with urbanisation, education, and intergroup contact (Lowe, 2021). If β is endogenous to the level of cross-caste interaction, then an initial reduction in trade barriers could lower β through increased exposure, generating a multiplier effect. Conversely, if social backlash raises β in response to integration, the short-run gains could partially reverse.

Technology and platforms. Digital marketplaces and e-commerce platforms may reduce the salience of the seller's identity in consumer transactions. If the taste-bias parameter β is lower for online transactions—because consumers cannot observe or do not care about the caste of the seller—then the expansion of digital platforms would constitute a de facto reduction in β. Our model can be used to quantify the real income gains from such a shift, though a full analysis would require modelling the adoption decision and the dual market structure (offline versus online).

8 Conclusion

This paper combines microdata on firms, workers, and consumers in India with a quantitative trade model to assess the aggregate and distributional costs of caste-based market segmentation. We document four facts establishing that cross-caste hiring is prevalent but asymmetric, tracks the caste composition of local consumer demand, and responds causally to caste-differentiated demand shocks. A central empirical finding is demand-led diversification: firms hire workers from other caste groups not despite identity barriers but because doing so allows them to reach consumers who would otherwise be inaccessible. The quantitative model embeds this mechanism in a general-equilibrium framework, nesting demand-side taste bias and supply-side hiring wedges with each friction identified from a distinct set of empirical moments.

Supply-side hiring frictions emerge as the dominant source of misallocation: removing hiring wedges raises aggregate real income by 4.7%, roughly twice the gain from eliminating taste bias. The two channels are complementary, and the gains accrue disproportionately to the most disadvantaged group, whose small home market makes cross-caste exchange most valuable. At 21.5% ad valorem on average, identity-based barriers within local markets are comparable in magnitude to the geographic frictions that separate Indian states—social distance can be as costly as physical distance. Labour-market interventions—anti-discrimination enforcement, standardised worker certification, cross-caste apprenticeships—complemented by market-access policies such as procurement set-asides and supply-chain inclusion mandates, are therefore likely to yield larger and more equitable gains than interventions aimed at consumer preferences alone.

More broadly, the framework developed here applies wherever social identity segments product markets—along ethnic, religious, racial, or sectarian lines. Quantifying the costs of identity-based segmentation in these contexts, and understanding whether demand-led diversification operates similarly, is a natural direction for future work.

References

AFRIDI, F., A. DHILLON, AND S. SHARMA (2024): "The ties that bind us: Social networks and productivity in the factory," *Journal of Economic Behavior & Organization*, 218, 470–485.

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.

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

ARROW, K. J. (1973): "The Theory of Discrimination," *Discrimination in Labor Markets*, 3–33.

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.

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.

BAU, N. AND A. MATRAY (2023): "Misallocation and capital market integration: Evidence from India," *Econometrica*, 91, 67–106.

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

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.

BERTRAND, M. AND E. DUFLO (2017): "Field Experiments on Discrimination," in *Handbook of Economic Field Experiments*, North-Holland, vol. 1, 309–393.

BHAGAVATULA, S., M. BHALLA, M. GOEL, AND B. VISSA (2023): "Social diversity in corporate boards and firm outcomes," *Journal of Corporate Finance*, 83, 102499.

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

BOHREN, J. A., P. HULL, AND A. IMAS (2022): "Systemic Discrimination: Theory and Measurement," Working Paper.

BOKEN, J., L. GADENNE, T. NANDI, AND M. SANTAMARIA (2025): "Community networks and trade," *Conditionally Accepted at the Journal of European Economic Association*.

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.

BORJAS, G. J. AND S. G. BRONARS (1989): "Consumer Discrimination and Self-Employment," *Journal of Political Economy*, 97, 581–605.

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

BÉTEILLE, A. (1965): *Caste, Class and Power: Changing Patterns of Stratification in a Tanjore Village*, Berkeley: University of California Press.

CASSAN, G., D. KENISTON, AND T. KLEINEBERG (2025): "A division of laborers: Identity and efficiency in India," Tech. rep., National Bureau of Economic Research.

CHARLES, K. K. AND J. GURYAN (2008): "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination," *Journal of Political Economy*, 116, 773–809.

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.

COOK, L. D., M. E. JONES, T. D. LOGAN, AND D. ROSÉ (2023): "The evolution of access to public accommodations in the United States," *The Quarterly Journal of Economics*, 138, 37–102.

——— (2025): "Competition and Consumer Discrimination," Tech. rep., National Bureau of Economic Research.

DE JANVRY, A., M. RAO, AND E. SADOULET (2025): "Seeding the seeds: Role of social structure in agricultural technology diffusion," *Journal of Economic Behavior & Organization*, 236, 107088.

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.

DUMONT, L. (1980): *Homo Hierarchicus: The Caste System and Its Implications*, Chicago: University of Chicago Press, complete revised english ed.

EDELMAN, B. G. AND M. LUCA (2014): "Digital discrimination: The case of Airbnb. com," *Harvard business school NOM unit working paper*.

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 (2025): "Cultural proximity and inter-firm trade," *European Economic Review*, 105129.

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

GIULIANO, L., D. I. LEVINE, AND J. LEONARD (2009): "Manager race and the race of new hires," *Journal of Labor Economics*, 27, 589–631.

GLOVER, D., A. PALLAIS, AND W. PARIENTE (2017): "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *Quarterly Journal of Economics*, 132, 1219–1260.

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

GOULD, H. A. (1964): "A Jajmani System of North India: Its Structure, Magnitude, and Meaning," *Ethnology*, 3, 12–41.

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.

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

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

HOLZER, H. J. AND K. R. IHLANFELDT (1998): "Customer discrimination and employment outcomes for minority workers," *The Quarterly Journal of Economics*, 113, 835–867.

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., E. HURST, C. I. JONES, AND P. J. KLENOW (2019): "The allocation of talent and us economic growth," *Econometrica*, 87, 1439–1474.

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.

HURST, E., A. Y. RUBINSTEIN, AND K. SHIMIZU (2024): "A Task-Based Framework for Understanding Racial Disparities," Working Paper.

IIDS (2010): "Dismantling Descent-Based Discrimination: Report on Dalits' Access to Rights," See the section reporting ActionAid survey statistics on consumer market discrimination (e.g., restrictions on Dalits buying/selling in markets and milk cooperatives).

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

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.

KLINE, P., E. K. ROSE, AND C. R. WALTERS (2022): "Systemic Discrimination Among Large U.S. Employers," *Quarterly Journal of Economics*, 137, 1963–2036.

KOLENDA, P. M. (1963): "Toward a Model of the Hindu Jajmani System," *Human Organization*, 22, 11–31.

LANG, K. AND J.-Y. K. LEHMANN (2012): "Racial Discrimination in the Labor Market: Theory and Empirics," *Journal of Economic Literature*, 50, 959–1006.

LIST, J. A. (2004): "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics*, 119, 49–89.

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.

MIAN, A. AND A. SUFI (2014): "What explains the 2007–2009 drop in employment?" *Econometrica*, 82, 2197–2223.

MINISTRY OF MICRO, SMALL AND MEDIUM ENTERPRISES (2024): "Annual Report 2024-25," Annual report, Government of India, reports MSME sector contribution: 30% of GDP, 35% of manufacturing output, 46% of total exports.

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.

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.

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

RAUCH, J. E. (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.

RESTUCCIA, D. AND R. ROGERSON (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 11, 707–720.

RUBINSTEIN, A. Y. (2025): "Is the Customer Always Right? Customer Taste-Based Discrimination and Competition," Working Paper, University of Pennsylvania.

SRINIVAS, M. N. (1962): *Caste in Modern India and Other Essays*, Bombay: Asia Publishing House.

——— (1994): *The Dominant Caste and Other Essays*, Oxford India Paperbacks, Delhi: Oxford University Press.

——— (2003): "An obituary on caste as a system," *Economic and political weekly*, 455–459.

STARTZ, M. (2025): "The value of face-to-face: Search and contracting problems in Nigerian trade," *Conditionally accepted at Econometrica*.

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*, 2, 256–281.

VAN LEEMPUT, E. (2021): "A passage to India: Quantifying internal and external barriers to trade," *Journal of International Economics*, 131, 103473.

WISER, W. H. (1936): *The Hindu Jajmani System*, Lucknow: Lucknow Publishing House.

Identity, Market Access, and Demand-led Diversification

Sampreet Singh Goraya and Akhil Ilango

Online Appendix

A Data

A.1 Caste Classification

India's caste system historically stratified society into endogamous groups with hereditary occupational roles. The Indian government classifies households into four broad social groups for the purposes of affirmative action and statistical enumeration: *Scheduled Castes* (SC), *Scheduled Tribes* (ST), *Other Backward Classes* (OBC), and *General* (upper castes). In both the NSS and MSME Census, the social group of a household is recorded in the household characteristics module as a self-reported variable. The NSS Schedule 10 records social group in Block 3 (Household Characteristics), Item 6, with codes: ST = 1, SC = 2, OBC = 3, Others = 9. For the MSME Census, the caste of the enterprise owner is recorded similarly. Following the literature, we aggregate these into three groups: *Lower Castes* (LC), comprising SC and ST households; *Middle Castes* (MC), comprising OBC households; and *Higher Castes* (HC), comprising General category households.

A.2 Micro, Small, and Medium enterprise Dataset

Main variables. The MSME data set spans non-agricultural enterprises and contains a representative sample of micro, small, and medium enterprises investing less than INR 100 million (manufacturing sector) or INR 50 million (services sector).

The MSME data set has two parts: a census of registered MSMEs and a sample survey of unregistered MSMEs. A total of 126,169 enterprises are surveyed to capture a representative sample of unregistered MSMEs. There are 1.65 million observations in total; we drop observations if any one of the revenues, capital stock, wage bill, or the number of employees is missing. Our empirical specification exploits rainfall shocks, that affect the rural economy, thus we restrict our attention to the rural areas. Following Jayachandran (2006), we also restrict our attention to 17 major states and union territories 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.

Table A.1 classifies sectors by caste concentration at two levels of aggregation. A sector is *caste-specialized* if more than 80% of firms are owned by a single caste group, and *overlapping* otherwise. At the 4-digit NIC level (Panel A), overlapping sectors account for 177 of 195 rural sectors, 99% of firms, and 98% of revenue. At the finer ASICC product code level (Panel B), overlapping products still account for 94% of rural firms and 75% of rural revenue, even though more products appear specialized when measured at this granular level. This concentration pattern implies that our demand-channel mechanism—where consumer identity preferences distort the allocation of demand across firms—operates in the vast majority of the economy.

Table A.1: Sector Classification by Caste Specialization

	Rural				Urban			
	Sectors	Firms (%)	Emp. (%)	Rev. (%)	Sectors	Firms (%)	Emp. (%)	Rev. (%)
Panel A: 4-Digit NIC sectors								
HC-specialized (>80% HC)	15	0.8	1.1	1.7	41	7.7	11.2	17.2
MC-specialized (>80% MC)	3	0.0	0.0	0.0	7	0.0	0.0	0.0
LC-specialized (>80% LC)	0	0.0	0.0	0.0	1	0.0	0.0	0.0
Overlapping (no group >80%)	177	99.2	98.9	98.3	153	92.3	88.8	82.8
Total	195	100	100	100	202	100	100	100
Panel B: ASICC product codes								
HC-specialized (>80% HC)	1,567	4.3	11.8	24.2	2,452	17.9	24.0	33.5
MC-specialized (>80% MC)	402	1.1	1.8	1.0	206	0.6	1.1	0.6
LC-specialized (>80% LC)	112	1.1	0.8	0.2	49	0.0	0.0	0.0
Overlapping (no group >80%)	2,509	93.6	85.6	74.6	2,774	81.6	74.9	65.9
Total	4,590	100	100	100	5,481	100	100	100

Notes. Panel A classifies sectors at the 4-digit NIC level; Panel B uses ASICC product codes. For each sector/product, we compute the weighted share of firms owned by HC, MC, and LC entrepreneurs using MSME sampling multipliers. A sector is classified as caste-specialized if more than 80% of its firms belong to a single caste group, and overlapping otherwise. Columns report the number of sectors, and the share of total weighted firms, employment, and revenue in each category. The sample includes all registered MSME firms (rural and urban). Source: MSME Census 2006–07.

Table A.2 lists the ten largest overlapping sectors ranked by their contribution to national MSME revenue. Together, these ten sectors account for over 37% of total revenue. All sectors feature meaningful caste heterogeneity in both firm ownership and employment, confirming that competitive interaction across caste groups is pervasive in the most economically important parts of the MSME sector.

A detailed example: Custom Tailoring (NIC 18105). To illustrate how overlapping sectors operate in practice, Table A.3 presents a detailed profile of custom tailoring (NIC 18105), the quintessential local service sector. Panel A shows that firm ownership is shared across all three caste groups: in rural areas, MC entrepreneurs own 53% of tailoring firms, HC own 27%, and LC own 19%. No single group dominates, making this a textbook overlapping sector. Panel B reports the employment composition within firms by owner caste. The diagonal entries confirm strong own-caste hiring—74% of workers in HC-owned rural tailoring firms are HC, 80% in MC-owned firms are MC, and 69% in LC-owned firms are LC—but the off-diagonal entries reveal substantial cross-caste employment in every cell. Panel C documents within-district overlap: of 430 rural districts with custom tailoring firms, 290 (67%) have firms owned by all three caste groups, and in 184 of these (63%) each group holds at least a 10% share. This pervasive local coexistence of tailors from different castes creates the competitive environment in which demand-side identity preferences can shape firm outcomes.

Table A.2: Top 10 Overlapping Sectors by Revenue Contribution

			Firm Ownership (%)			Employment (%)		
NIC	Sector Description	Rev. (%)	HC	MC	LC	HC	MC	LC
Panel A: Rural								
1531	Grain mill products	13.3	49.0	42.9	8.1	39.0	40.5	20.4
1711	Spinning & weaving of textiles	8.0	50.8	38.9	10.3	34.4	39.3	26.3
1514	Vegetable & animal oils	5.4	55.9	39.7	4.4	45.3	39.7	15.0
1542	Sugar	4.1	44.9	47.3	7.7	37.9	38.2	23.9
2714	Basic iron & steel	3.4	75.8	18.7	5.5	54.5	19.4	26.1
2715	Ferro-alloys	2.6	79.2	15.8	5.0	52.8	25.1	22.2
1520	Dairy products	2.5	56.0	36.9	7.2	55.2	31.4	13.4
2520	Plastics products	2.5	66.1	22.0	11.9	53.8	27.2	19.0
2899	Other fabricated metal products	2.5	40.6	53.8	5.6	44.7	39.6	15.7
1810	Wearing apparel	2.4	29.3	52.1	18.5	40.7	42.9	16.4
Panel B: Urban								
1531	Grain mill products	6.4	53.0	42.2	4.8	42.6	38.6	18.8
1810	Wearing apparel	6.1	38.9	52.4	8.7	46.3	37.1	16.6
2892	Metal treatment & machining	4.2	61.3	34.0	4.7	64.8	25.3	9.9
1711	Spinning & weaving of textiles	4.1	78.1	18.5	3.5	52.9	30.3	16.8
2899	Other fabricated metal products	3.6	71.9	24.2	3.9	62.8	24.1	13.1
1514	Vegetable & animal oils	3.0	60.7	36.6	2.7	47.3	34.7	18.0
2520	Plastics products	2.9	76.8	18.9	4.3	63.9	23.1	13.0
1712	Finishing of textiles	2.5	67.3	28.0	4.6	53.0	30.1	16.9
3591	Motorcycles & parts	1.9	74.9	18.0	7.1	71.0	17.6	11.5
1549	Other food products n.e.c.	1.8	60.9	35.3	3.9	46.6	35.0	18.4

Notes. The table reports the ten overlapping sectors (no single caste group >80% of firms) with the largest share of national MSME revenue. Rev. (%) is each sector's share of total weighted revenue across all sectors. Firm Ownership columns report the percentage of weighted firms owned by each caste group. Employment columns report the percentage of weighted total employment accounted for by workers of each caste. The sample includes all registered MSME firms (rural and urban). Source: MSME Census 2006–07.

Table A.3: Overlapping Sector Example: Custom Tailoring (NIC 18105)

Panel A: Firm Ownership

	HC (%)	MC (%)	LC (%)	Total Firms
Rural	27.4	53.4	19.2	57,463
Urban	33.9	56.8	9.3	66,314

Panel B: Employment Composition (%)

	Rural			Urban		
	HC-owned	MC-owned	LC-owned	HC-owned	MC-owned	LC-owned
HC workers	73.8	14.5	19.0	83.3	18.5	37.2
MC workers	18.4	80.1	12.2	12.8	75.9	18.6
LC workers	7.8	5.4	68.8	3.8	5.7	44.2

Panel C: Within-District Overlap (Rural)

Districts with tailoring firms	430
Districts with all 3 castes present	290 (67.4%)
Districts with each caste ≥10%	184 (63.4%)
Mean district share: HC / MC / LC	31.0 / 47.8 / 21.2

Notes. Custom tailoring is identified by 5-digit NIC activity code 18105 within the broader "Manufacture of wearing apparel" division (NIC 1810). Panel A reports the weighted share of firms owned by each caste group and the total number of weighted firms. Panel B reports the share of total employment by worker caste (rows) within firms grouped by owner caste (columns), separately for rural and urban areas. Panel C examines within-district overlap in rural India: the number of districts containing custom tailoring firms from multiple caste groups, and the share where each group holds at least 10% of local firms. Mean district shares are computed across the 290 districts with all three castes present. Source: MSME Census 2006–07. Sampling multipliers are applied.

Table A.4: Population Shares and Consumption Expenditure by Caste Group (2006)

	Rural			Urban			Overall		
	HC	MC	LC	HC	MC	LC	HC	MC	LC
Panel A: Shares (%)									
Household share	23.5	42.1	34.4	46.7	35.7	17.6	29.7	40.4	29.9
Population share	23.5	43.2	33.3	45.7	36.6	17.8	28.9	41.6	29.5
Expenditure share	28.2	44.0	27.8	55.9	30.7	13.3	38.8	38.9	22.3
Panel B: Monthly per-capita expenditure (Rs.)									
MPCE	721	602	493	1531	1043	903	1062	706	557
MPCE (HC = 1)	1.00	0.83	0.68	1.00	0.68	0.59	1.00	0.66	0.52
Panel C: Household characteristics									
Household size	4.93	5.05	4.75	4.23	4.44	4.37	4.64	4.91	4.69
Literacy rate (%)	69.9	55.1	43.3	88.3	80.5	69.0	77.6	61.1	47.3
Years of education	5.5	3.8	2.9	9.2	7.0	5.6	7.1	4.6	3.3
Panel D: Individual characteristics (NSS Employment Survey)									
Employed in agriculture (%)	26.0	47.0	59.0	–	–	–	26.0	47.0	59.0

Notes. Data from the NSS 62nd Round (2005–06) for Panels A–C; NSS Employment and Unemployment Survey (2004–2010) for Panel D. HC = General (upper castes); MC = OBC (Other Backward Classes); LC = SC/ST (Scheduled Castes and Scheduled Tribes). Household and population shares are computed using NSS sampling multipliers; population shares weight by household size. Expenditure shares use total household expenditure (MPCE $\times$ household size $\times$ weight). MPCE is CPI-deflated monthly per-capita expenditure in rupees. The normalized row expresses each group's MPCE relative to HC = 1. Literacy rate is the weighted share of household heads classified as literate (general education code $\geq$ 2) in the NSS person-level records (Block 4, Item 7). Years of education maps the NSS general education codes to approximate completed years: not literate = 0, literate without formal schooling = 2, below primary = 3, primary = 5, middle = 8, secondary = 10, higher secondary/diploma = 12, graduate = 15, postgraduate and above = 17. Panel D reports individual-level employment statistics; urban-rural breakdown not available in the employment survey. All statistics use NSS sampling weights.

A.3 Household Consumption Data

We use four schedules spanning the years 2003–04 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 A.4 (Panels A–C).

Table A.5 presents the consumption group categories used in our analysis. The NSS consumption data records household expenditure across granular subcategories (Column 3), which we aggregate into 16 medium categories (Column 2) for the demand–employment regressions (Fact 3), and further into 8 broad categories (Column 1) for the consumption elasticity regressions. The first four medium categories—cereals & pulses, dairy/oil/sugar, fruits/vegetables/meat/spices, and beverages & prepared food—are classified as food products in the food interaction specification. The table also reports the weighted number of registered MSME firms matched to each category and the share of firms owned by each caste group, separately for rural and urban areas. Approximately 13% of rural and 17% of urban firms are not matched to any NSS consumption category.

Table A.5: Consumption Group Categories and MSME Firm Distribution

			Rural				Urban			
Broad (8)	Medium (16)	Granular NSS Subcategories	Firms	HC	MC	LC	Firms	HC	MC	LC
Food	Cereals & pulses	Cereals; Pulses & products	21,212	56.3	37.4	6.3	12,242	63.3	32.4	4.3
	Dairy, oil & sugar	Milk & milk products; Edible oil; Sugar; Salt	12,758	57.5	37.0	5.5	10,711	63.4	32.8	3.8
	Fruits, veg, meat & spices	Egg, fish & meat; Vegetables; Fruits (fresh); Fruits (dry); Spices	7,701	65.9	28.9	5.3	5,759	59.4	36.6	4.0
	Beverages & prepared food	Beverages etc.; Food (prepared)	53,921	46.7	45.0	8.3	31,599	53.2	42.1	4.7
Tobacco & intox	Tobacco & intoxicants	Pan; Tobacco; Intoxicants	4,156	61.8	30.5	7.8	5,340	68.2	28.1	3.7
Fuel	Fuel & light	Fuel and light	4,380	31.8	62.2	6.0	4,533	47.3	47.9	4.8
Clothing	Clothing & bedding	Clothing; Bedding	32,528	45.2	40.1	14.7	82,982	70.2	24.8	5.0
Footwear	Footwear	Footwear	13,898	35.5	18.7	45.8	32,728	60.2	20.2	19.6
Medical	Medical	Medical (non-institutional); Therapeutic appliances	21,981	42.4	49.0	8.7	55,645	61.8	33.4	4.8
Misc	Services	Entertainment; Consumer services; Conveyance	140,448	35.8	47.0	17.2	212,850	48.7	42.8	8.6
	Toilet & sundry	Personal effects; Toilet articles; Sundry articles	8,781	44.2	33.1	22.7	8,764	66.1	25.8	8.2
Durables	Furniture	Furniture & wood products	38,801	26.7	59.3	14.1	36,112	50.8	43.0	6.2
	Electronics & appliances	Electronics (recreation); Cooking appliances; Other personal goods	6,521	64.2	28.9	6.9	21,935	77.3	18.8	3.9
	Jewellery	Jewellery & ornaments	2,290	42.9	42.9	14.2	7,440	76.2	21.0	2.8
	Crockery & utensils	Crockery & utensils	19,824	46.4	46.9	6.7	44,426	67.4	28.3	4.3
	Transport	Transport equipment	3,135	55.9	34.5	9.7	13,173	76.0	19.2	4.8
	Not matched		56,970	49.1	40.3	10.5	116,775	69.8	25.6	4.6
	Total		449,305	42.5	44.0	13.5	703,013	60.4	33.0	6.6

Notes. Column 1 lists the 8 broad consumption categories used in the consumption elasticity regressions. Column 2 lists the 16 medium categories used in the demand–employment regressions (Fact 3). Column 3 lists the corresponding granular NSS subcategories from NSS Round 61 (2004–05). "Firms" reports the weighted number of registered MSME firms matched to each category. HC, MC, and LC columns report the percentage of firms owned by each caste group. "Not matched" includes firms whose ASICC product code could not be mapped to an NSS consumption category. Source: MSME Census 2006–07.

A.4 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.[12] Table A.4 (Panel D) shows that LC individuals are more than twice as likely to be employed in agriculture (59%) compared to HC individuals (26%).

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

A.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 percentile of districts with excessive rainfall to avoid cases of floods. For the analysis, we define rain shock as equal to +1 for positive shock, -1 for negative shock, and 0 otherwise.

A.6 Urban Firm Distribution

In Table A.6, we provide the firm-size distribution for urban registered MSME firms; the average firm size is 6.9 employees and the majority (3.9) of the employees belong to the HC community and 1.0 employee is from the LC community.

Table A.6: Firm-size distribution: Urban firms

	All Firms						Mean by Owner Caste		
	Mean	Median	p5	p95	N	Firms (wt.)	HC (424,635)	MC (231,956)	LC (46,422)
Emp. All	6.9	3	1	20	691,169	703,013	8.2	4.8	5.5
Emp. LC	1.0	0	0	4	691,169	703,013	1.1	0.7	1.8
Emp. MC	2.0	1	0	6	691,169	703,013	1.6	2.9	1.2
Emp. HC	3.9	2	0	11	691,169	703,013	5.5	1.3	2.6
Emp. Own-C (%)	74	100	0	100	691,165	703,013	77	74	48
Revenue (10^4)	535	28.2	4.3	999	691,169	703,013	738	211	326
Materials (10^4)	327	17.8	0.5	654	691,169	703,013	457	120	192

Notes. Table A.6 presents the firm size distribution of urban registered MSME firms. 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. N is the unweighted number of firms in the sample; Firms (wt.) is the weighted count using MSME sampling multipliers. The last three columns report the mean for firms owned by each caste group; the number of weighted firms in each group is shown in parentheses below the column header. Source: MSME 2006–07.

A.6.1 Fact 1

In Table A.7, we report the worker composition among urban firms.

Table A.7: Cross-Caste Hiring Patterns: Alternative Samples

	Employment Shares (%)			Extensive Margin (%)		
	HC-owned	MC-owned	LC-owned	HC-owned	MC-owned	LC-owned
Panel A: Urban, all sectors						
HC workers	67.4	26.2	47.5	92.1	30.4	43.2
MC workers	19.2	59.3	21.0	33.0	86.5	29.4
LC workers	13.4	14.5	31.5	20.2	16.3	58.5
Random benchmark	HC: 57.0 / MC: 28.4 / LC: 14.6					
Panel B: Urban, overlapping sectors						
HC workers	66.9	25.4	46.2	91.7	29.8	42.1
MC workers	19.7	60.1	21.1	33.3	86.7	29.0
LC workers	13.5	14.5	32.7	20.0	16.0	59.1
Random benchmark	HC: 55.6 / MC: 29.6 / LC: 14.8					

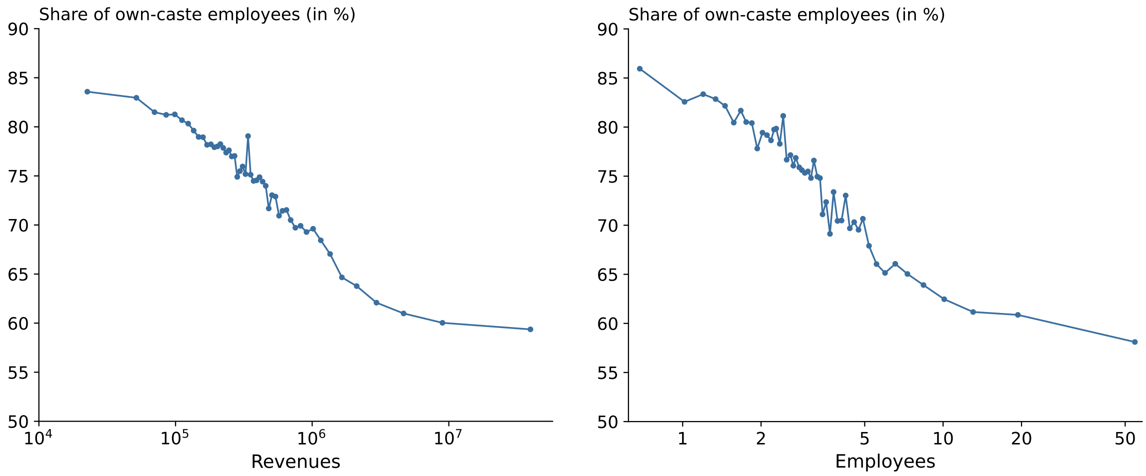
Notes. Employment Shares report the share of total employment by worker caste (rows) within firms grouped by owner caste (columns). Extensive Margin reports the share of firms (by owner caste) that employ at least one worker of the indicated caste. The *Random benchmark* row reports the overall employment composition across all firms in the subsample; under random hiring (proportional to the labor pool), every owner-caste column would equal this benchmark. Panel A includes all rural registered MSME firms. Panel B restricts to rural firms in overlapping sectors (no single caste group owns more than 80% of firms at the 4-digit NIC level). Panel C includes all urban firms. Panel D restricts to urban firms in overlapping sectors. Source: MSME 2006–07. Sampling multipliers are applied.

A.6.2 Fact 2

This section provides robustness checks for the negative relationship between firm size and own-caste employment documented in Section 4.2. Figure A.1 replicates the baseline analysis for urban firms, showing that the steep decline in own-caste share as firms grow holds outside rural areas. The pattern is quantitatively similar to the rural baseline (Figure 2): own-caste employment falls from approximately 85% among the smallest urban firms to 60% among the largest, whether measured by revenue or total employment.

Figure A.2 uses an alternative measure of workforce diversity: the caste Herfindahl–Hirschman Index (HHI), which ranges from 1 (perfectly homogeneous workforce) to 1/3 (equal representation of all three caste groups). The HHI declines sharply with firm size in both rural and urban areas, confirming that larger firms hire more diverse workforces. The negative relationship is robust to this alternative specification, ruling out concerns that the baseline results are driven by the particular definition of own-caste share.

Figure A.3 decomposes the workforce composition by owner caste, plotting the employment shares of LC, MC, and HC workers separately for firms owned by each caste group. For all three owner castes, larger firms employ more workers from other caste groups. LC-owned firms exhibit the steepest decline in own-caste share, consistent with the demand-driven diversification mechanism: as LC firms grow, they hire workers from higher castes to reach broader consumer markets. Online Appendix Table A.8



(a) Firm size measured by revenue

(b) Firm size measured by employment

Figure A.1: Homophily in Hiring and Firm Size: Urban Firms

Notes. Figure A.1a presents a bin-scatter plot for urban firms; the x-axis is the total revenues, and the y-axis is the share of own-caste workers. Figure A.1b presents a bin-scatter plot for urban firms; the x-axis is the total employees, and the y-axis is the share of own-caste workers. We control for caste, district and 4-digit sector fixed effects. The sample consists of 691,165 registered urban MSME firms. Source: MSME 2006–07. Sampling multipliers are applied.

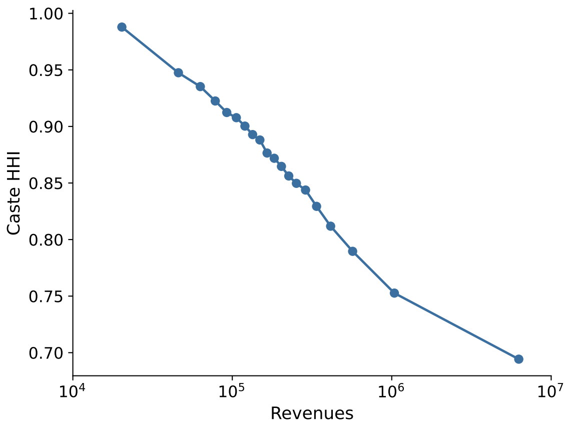
presents regression estimates quantifying these relationships, including interactions with contact-intensive sectors where employee–consumer interaction is salient.

A.6.3 Fact 3

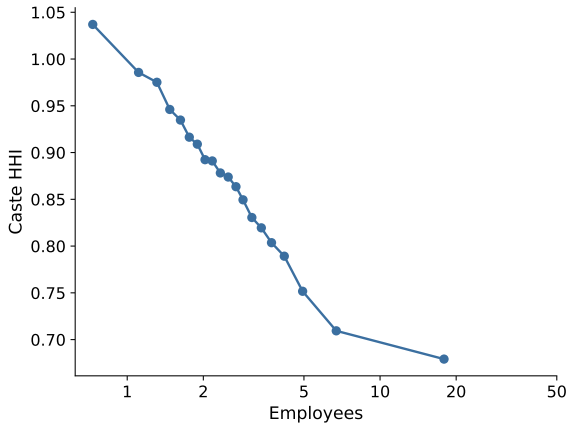
This section provides details on the cross-sectional demand–employment regressions described in Section 4.3. The demand shares are constructed from NSS Round 61 (2004–05) household consumption data. For each district d, product category p, and sector (rural/urban), we compute the weighted consumption expenditure by caste group:

$$\text{DemandShare}^s_{dp} = \frac{\sum_{h\in s} w_h \cdot x_{hdp}}{\sum_h w_h \cdot x_{hdp}},$$

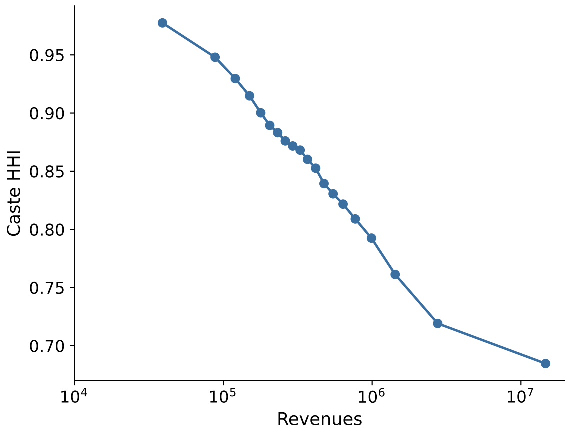
where w_h is the NSS sampling weight and x_{hdp} is the annual expenditure of household h on product p in district d. The MSME product codes (ASICC) are mapped to NSS consumption categories using a bottom-up concordance that aggregates fine product codes into 16 medium categories: cereals & pulses, dairy/oil/sugar, fruits/vegetables/meat/spices, beverages & prepared food, tobacco & intoxicants, fuel & light, clothing & bedding, footwear, medical, services, toilet & sundry, furniture, electronics & appliances, jewellery, crockery & utensils, and transport. The food indicator in Equation (2) takes value one for the first four categories. The sample includes 333,639–363,525 firm-level observations (depending on the regressor caste group), covering registered MSME firms in rural districts.



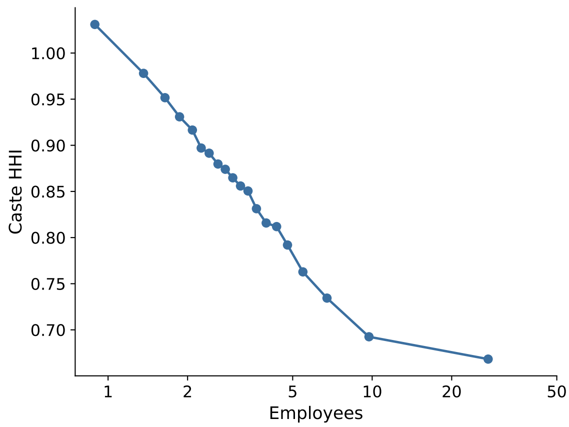
(a) Rural: Firm size by revenue



(b) Rural: Firm size by employment



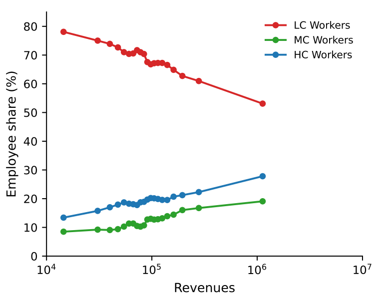
(c) Urban: Firm size by revenue



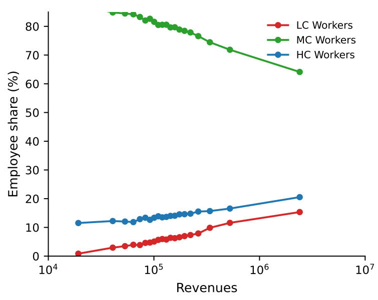
(d) Urban: Firm size by employment

Figure A.2: Caste HHI of Employees and Firm Size

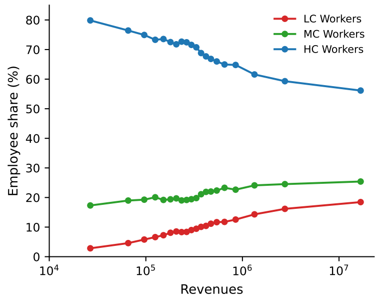
Notes. Each panel plots the caste Herfindahl–Hirschman Index (HHI) of a firm's workforce against firm size, using 50-quantile binscatter. HHI $= \sum_s \omega_s^2$, where ω_s is the share of caste s in total employment. HHI $= 1$ indicates a perfectly homogeneous workforce; HHI $= 1/3$ indicates equal representation. We control for caste, district and 4-digit sector fixed effects. Panels (a) and (b) show rural registered MSME firms (N = 444,574); panels (c) and (d) show urban registered MSME firms (N = 691,165). Source: MSME 2006–07. Sampling multipliers are applied.



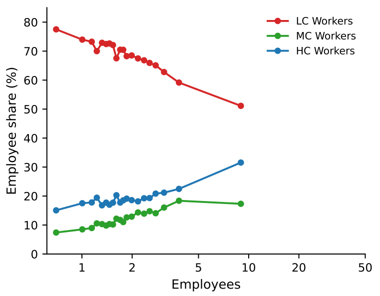
(a) Rural Rev: LC-Owned



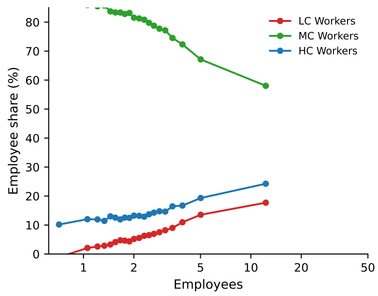
(b) Rural Rev: MC-Owned



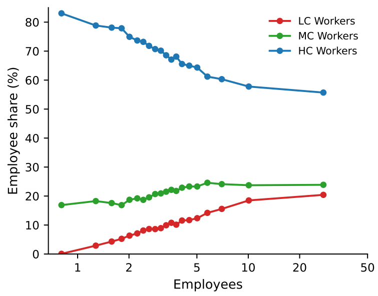
(c) Rural Rev: HC-Owned



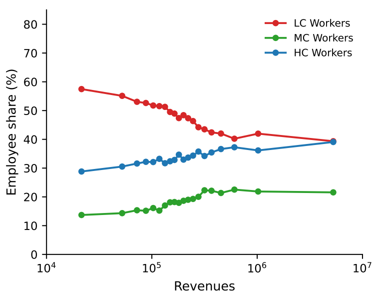
(d) Rural Emp: LC-Owned



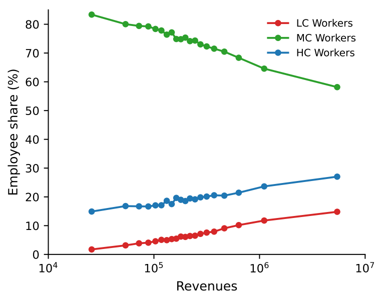
(e) Rural Emp: MC-Owned



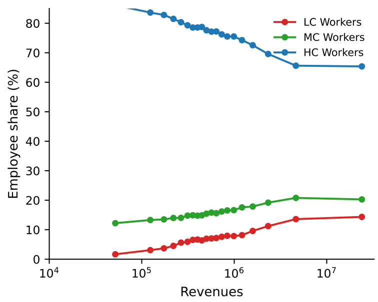
(f) Rural Emp: HC-Owned



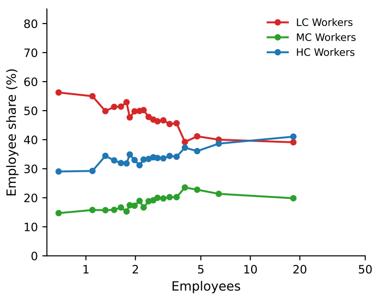
(g) Urban Rev: LC-Owned



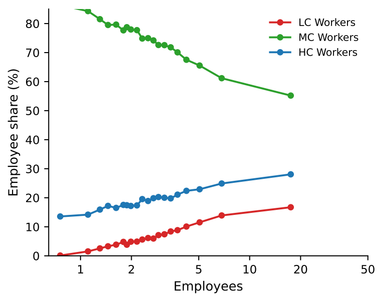
(h) Urban Rev: MC-Owned



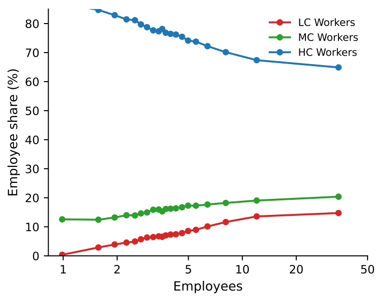
(i) Urban Rev: HC-Owned



(j) Urban Emp: LC-Owned



(k) Urban Emp: MC-Owned



(l) Urban Emp: HC-Owned

Figure A.3: Employee Caste Composition by Owner Caste and Firm Size

Notes. Each panel plots the share of LC, MC, and HC employees against firm size (50-quantile binscatter), separately by owner caste. We control for district and 4-digit sector fixed effects. Rows 1-2 use rural registered MSME firms ($N = 444{,}574$); rows 3-4 use urban registered MSME firms ($N = 691{,}165$). Rows 1 and 3 measure firm size by revenue; rows 2 and 4 measure firm size by number of employees. Source: MSME 2006–07. Sampling multipliers are applied.

Table A.8: Own-Caste Employee Share and Firm Size

	Dep. var.: Own-caste employee share (%)			
	(1)	(2)	(3)	(4)
log(Employees)	-10.193^{***}	-10.101^{***}	-9.006^{***}	-9.000^{***}
	(0.423)	(0.426)	(0.560)	(0.565)
log(Employees) × Contact		-1.010^{*}		1.448
		(0.518)		(0.778)
log(Employees) × MC			-2.301^{***}	-2.071^{**}
			(0.786)	(0.781)
log(Employees) × LC			-4.069^{***}	-3.818^{***}
			(1.014)	(1.044)
log(Employees) × MC × Contact				-4.814^{***}
				(1.359)
log(Employees) × LC × Contact				-3.826^{**}
				(1.746)
Observations	440,642	440,642	440,642	440,642
Adjusted R-squared	0.182	0.182	0.183	0.183
District FE	✓	✓	✓	✓
NIC-4 FE	✓	✓	✓	✓
Caste FE	✓	✓	✓	✓

Notes. The dependent variable is the own-caste employee share (%). Column (1) includes log employment as the only regressor. Column (2) adds an interaction with contact-intensive sectors. Column (3) adds interactions with owner caste (MC, LC). Column (4) is the full model with triple interactions. "Contact" is a dummy for contact-intensive sectors (carpentry and furniture manufacturing, construction, wholesale and retail, hotels and restaurants, travel agencies, post and telecommunications, and computer-related services). LC and MC are owner-caste dummies (HC is the omitted category). All specifications include caste, district, and NIC-4 sector fixed effects. Sample: rural registered MSME firms. Sampling multipliers are applied. Standard errors clustered at district level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.6.3.1 District-level controls. Panels D and E add district-level controls to the food interaction specification to absorb observable differences across districts that may confound the demand–employment relationship. All controls are constructed from NSS Round 63 (2006–07) household survey data. *Urbanization* is the share of sample households located in urban areas (sector = 2). *Agricultural share* is the fraction of households whose primary economic activity falls under NIC divisions 01–02 (agriculture and allied activities), identified from the household-level industry code in Block 3 of the survey schedule. *Literacy rate* is the share of household members aged 5 and above with an education level of "literate without formal schooling" or higher (general education code ≥ 2 in Block 4 of the person records). *Caste isolation index* measures the degree of residential segregation of caste groups within each district. Using the First Stage Unit (FSU) as a proxy for village/neighborhood, we compute the isolation index for each caste group s in district d as:

$$I_d^s = \sum_v \frac{n_v^s}{N_d^s} \cdot \frac{n_v^s}{n_v},$$

where n_v^s is the weighted count of caste-s households in FSU v, n_v is the total weighted household count in v, and N_d^s is the total caste-s population in district d. The index measures the probability that a randomly chosen neighbour of a caste-s household belongs to the same caste group, ranging from 0 (perfect integration) to 1 (complete isolation). We average the isolation indices across the three caste groups (HC, MC, LC) to obtain a single district-level measure.

A.6.4 Fact 4.1: Income, Wage and Consumption Elasticity

Table A.10 decomposes the baseline consumption elasticity into positive and negative rainfall shocks. Column (1) reproduces the benchmark from Table 3. Column (2) re-estimates the baseline specification on the subsample of district-years with non-negative rainfall shocks, and Column (3) on district-years with non-positive shocks; zero-shock districts serve as the control group in both cases. The LC consumption response is significant and substantial on both sides: a 20.5% increase for positive shocks and a 17.5% decrease for negative shocks (both relative to HC). The broadly symmetric pattern confirms that the baseline estimate is not driven by one tail of the rainfall distribution.

A.6.4.1 Wage and income elasticity. Table A.11 examines how individual wages respond to rainfall shocks, using the NSS Employment–Unemployment Schedule (Rounds 60–64, covering 2004–2008, and Round 66, covering 2010). The unit of observation is

Table A.9: Cross-Caste Demand–Employment Matrices: Rural and Urban

	All Products			Food Only			Non-Food Only		
Dep. var:	HC	MC	LC	HC	MC	LC	HC	MC	LC
Panel A: Rural									
HC emp. share	0.108***	-0.087***	-0.007	0.199***	-0.132***	-0.042	0.089***	-0.078**	0.007
	(0.033)	(0.031)	(0.037)	(0.055)	(0.050)	(0.063)	(0.032)	(0.032)	(0.035)
MC emp. share	-0.090***	0.120***	-0.072**	-0.125**	0.174***	-0.089	-0.080***	0.106***	-0.066*
	(0.028)	(0.029)	(0.036)	(0.054)	(0.050)	(0.061)	(0.027)	(0.028)	(0.035)
LC emp. share	-0.018	-0.033*	0.079***	-0.074***	-0.042*	0.131***	-0.008	-0.028	0.059**
	(0.023)	(0.020)	(0.021)	(0.022)	(0.024)	(0.028)	(0.026)	(0.022)	(0.024)
Panel B: Urban									
HC emp. share	0.147***	-0.169***	-0.041	0.211***	-0.200***	-0.030	0.143***	-0.165***	-0.044
	(0.037)	(0.047)	(0.066)	(0.050)	(0.047)	(0.072)	(0.038)	(0.049)	(0.068)
MC emp. share	-0.114***	0.165***	-0.005	-0.140***	0.180***	-0.055	-0.113***	0.164***	0.001
	(0.030)	(0.039)	(0.054)	(0.045)	(0.044)	(0.066)	(0.031)	(0.040)	(0.055)
LC emp. share	-0.033*	0.004	0.046*	-0.071***	0.020	0.085*	-0.030*	0.002	0.043*
	(0.017)	(0.019)	(0.025)	(0.027)	(0.032)	(0.043)	(0.017)	(0.018)	(0.026)

Panel C: Urban, food interaction

	Baseline			Food × Demand		
	HC	MC	LC	HC	MC	LC
HC emp. share	0.146***	-0.168***	-0.044	0.017	-0.011	0.037
	(0.037)	(0.047)	(0.068)	(0.030)	(0.026)	(0.066)
LC emp. share	-0.033*	0.006	0.043*	0.008	-0.024	0.039
	(0.017)	(0.019)	(0.026)	(0.018)	(0.020)	(0.048)
MC emp. share	-0.112***	0.162***	0.002	-0.025	0.035	-0.076
	(0.030)	(0.039)	(0.056)	(0.033)	(0.030)	(0.057)

Panel D: Rural, food interaction + district controls

	Baseline			Food × Demand		
	HC	MC	LC	HC	MC	LC
HC emp. share	0.093***	-0.071**	-0.003	0.068*	-0.029	-0.056
	(0.035)	(0.032)	(0.035)	(0.038)	(0.037)	(0.053)
MC emp. share	-0.084***	0.094***	-0.050	-0.040	0.071**	-0.035
	(0.028)	(0.028)	(0.034)	(0.040)	(0.034)	(0.047)
LC emp. share	-0.008	-0.023	0.053**	-0.027	-0.043*	0.091***
	(0.026)	(0.021)	(0.024)	(0.031)	(0.025)	(0.034)

Panel E: Urban, food interaction + district controls

	Baseline			Food × Demand		
	HC	MC	LC	HC	MC	LC
HC emp. share	0.116***	-0.123***	-0.028	0.010	-0.009	0.026
	(0.037)	(0.047)	(0.061)	(0.029)	(0.025)	(0.057)
MC emp. share	-0.084***	0.122***	-0.010	-0.019	0.032	-0.067
	(0.031)	(0.039)	(0.049)	(0.031)	(0.028)	(0.054)
LC emp. share	-0.032*	0.001	0.037	0.008	-0.024	0.041
	(0.016)	(0.018)	(0.025)	(0.018)	(0.020)	(0.048)

Notes. Each cell reports $\hat{\gamma}_{s \to s'}$ from Equation (1), with standard errors clustered at the district level in parentheses. Rows index the caste of employees (s'); columns index the caste whose demand share is used as the regressor (s). All regressions include state and product fixed effects. Panel A uses the rural sample; Panel B uses the urban sample. Within each area, the full sample is split into food products (cereals & pulses, dairy/oil/sugar, fruits/veg/meat/spices, beverages & prepared food) and non-food products. Panel C reports the urban food interaction specification from Equation (2): the baseline coefficient $\hat{\gamma}_{s \to s'}$ and the food × demand interaction term $\hat{\delta}_{s \to s'}$. Panels D and E repeat the food interaction specification for the rural and urban samples, respectively, adding district-level controls: urbanization rate, agricultural household share, average literacy rate, and a caste isolation index (computed at the FSU/village level within each district). All district controls are constructed from NSS Round 63 (2006–07). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.10: Consumption Elasticity: Positive vs. Negative Rainfall Shocks

	(1) Benchmark	(2) Positive	(3) Negative
Rainfall × MC	0.045**	0.087***	-0.085**
	(0.021)	(0.028)	(0.034)
Rainfall × LC	0.092***	0.205***	-0.175***
	(0.028)	(0.029)	(0.039)
Observations	111,312	91,596	93,538
R^2	0.275	0.297	0.321
Controls	✓	✓	✓
Caste FE	✓	✓	✓
District × Year FE	✓	✓	✓

Notes. The dependent variable is log monthly per-capita expenditure (MPCE). Column (1) reproduces the benchmark specification from Table 3, Column 1. Column (2) restricts to district-years with rainfall shock $\geq$ 0 and Column (3) to district-years with rainfall shock $\leq$ 0, each running the baseline specification on the respective subsample. All specifications include household-head controls (meals per day, education, land ownership) interacted with the rainfall shock. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

an individual worker i in district d in year t. We estimate:

$$\ln w_{idt} = \beta_s \cdot \text{Rshock}_{dt} \times \mathbf{1}[\text{caste} = s] + \mathbf{X}'_{idt}\gamma + \alpha_d + \alpha_t + \varepsilon_{idt},$$

where w_{idt} is either the daily wage or the total wage (daily wage × days worked) of individual i, both winsorized at the 1st and 99th percentiles. Controls $\mathbf{X}_{idt}$ include age, age squared, education, sex, land ownership, and month-of-survey dummies. All regressions include district and year fixed effects and are weighted by NSS sampling multipliers; standard errors are clustered at the district level.

Columns (1)–(3) use log daily wage as the outcome. The wage response to rainfall shocks is concentrated among agricultural workers (column 2): both MC and LC agricultural wages rise by approximately 5.7% relative to HC wages following a positive rainfall shock. Non-agricultural wages (column 3) show no significant response. Columns (4)–(6) use log total wage as the outcome, which captures both the daily wage rate and the number of days worked. The agricultural LC coefficient increases to 6.5% (column 5), suggesting that positive rainfall shocks raise LC agricultural earnings through both higher daily wages and more days of employment. The sample is larger for total wages because it includes workers for whom days worked are observed but daily wages are imputed or recorded differently.

Table A.11: Wage and Income Elasticity to Rainfall Shocks

	Log Daily Wage			Log Total Wage		
	(1) All	(2) Agri.	(3) Non-Agri.	(4) All	(5) Agri.	(6) Non-Agri.
Rainfall × MC	0.023	0.057**	-0.008	0.028	0.053**	0.002
	(0.021)	(0.026)	(0.020)	(0.020)	(0.025)	(0.019)
Rainfall × LC	0.013	0.057**	-0.007	0.024	0.065***	0.009
	(0.026)	(0.025)	(0.021)	(0.024)	(0.024)	(0.020)
Observations	192,425	73,003	119,422	237,105	85,839	151,266
R^2	0.493	0.322	0.492	0.492	0.308	0.488
Controls	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Notes. The sample consists of individual workers from the NSS Employment–Unemployment Schedule, Rounds 60–64 (2004–2008) and Round 66 (2010), restricted to rural areas. Columns (1)–(3) use log daily wage (winsorized at the 1st and 99th percentiles) as the dependent variable; Columns (4)–(6) use log total wage (daily wage × days worked, winsorized). Columns (1) and (4) include all workers; Columns (2) and (5) restrict to agricultural workers (`worktype` = 1); Columns (3) and (6) restrict to non-agricultural workers. Controls include age, age^2, education, sex, land ownership, and month-of-survey dummies. The rainfall shock is the deviation of rainfall from the historical district median, normalized by the median. NSS sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.6.5 Fact 4.2: Revenue Elasticity

Table A.12 examines how firm-level revenues respond to local demand shocks, and whether this response differs across caste groups.

A.6.5.1 Baseline results. Columns (1)–(2) estimate the regression on the full sample of firms. Both MC and LC firms show positive and significant revenue responses to rainfall shocks. The point estimates imply that a one-standard-deviation positive rainfall shock raises the revenues of LC-owned firms by approximately 11%, and MC-owned firms by approximately 9%, relative to HC-owned firms. These estimates are stable across the two fixed-effect specifications.

A.6.5.2 Product market segmentation. Columns (4)–(5) split the sample by whether the firm's primary product belongs to a low-caste expenditure (LCE) consumption category—food, fuel, and tobacco—or to other categories including durables, services, and intermediate goods. If the demand channel operates through local consumption, we expect the effect to be concentrated among products that LC households disproportionately consume. Column (4) shows that the LCE subsample exhibits a positive but imprecisely estimated LC coefficient, while column (5) shows that the remaining products—which include a broader set of goods traded across caste groups—exhibit strong and significant effects for both MC and LC firms. Column (6) restricts to firms producing non-tradable goods. We measure tradability using the index of Mian and Sufi (2014), which captures geographic concentration of industry employment; we construct a concordance from NAICS to NIC codes to apply this measure to Indian manufacturing products. Products with a tradability index below 0.6 are classified as non-tradable; this threshold is close to the sample mean, and the results are not sensitive to the choice of cutoff. Non-tradable firms, whose revenues depend more on local demand conditions, show strong and significant responses for both caste groups.

A.6.5.3 Price dispersion. A potential confound is that LC and HC firms may face different prices for the same product, in which case the revenue response could reflect price discrimination rather than demand. Columns (7)–(8) restrict to products with low output price dispersion across firms. Column (7) retains products where the absolute difference in mean log output price between LC and HC producers is below the 75th percentile, and column (8) retains products where the within-product standard deviation of log output price is below the 75th percentile. In both cases, the LC and MC revenue responses remain positive and significant, suggesting that the results are not driven by caste-based price differentials.

An analogous concern applies to the input side: if LC firms pay systematically different prices for raw materials, cost differences could generate differential revenue responses. Columns (10)–(11) address this by restricting to products where raw material price dispersion is low, using the same two measures (absolute LC–HC price gap and within-product SD) applied to log input prices. The results remain significant in both subsamples, confirming that the revenue response is not driven by differential input costs.

A.6.5.4 District-level segregation. Column (9) restricts to districts where caste groups are residentially integrated. We construct an isolation index from the 2001 Census:

$$I_d = \sum_v \frac{LC_v}{LC_d} \cdot \frac{LC_v}{T_v},$$

where LC_v and T_v are the SC/ST and total populations in village v, and LC_d is the total SC/ST population in district d. The index measures the extent to which LC households are concentrated in predominantly LC villages. A higher value indicates greater residential segregation: LC households live in areas with few non-LC neighbors, limiting the scope for cross-caste demand spillovers. A lower value indicates that LC households are more spatially integrated with other caste groups, so that local demand shocks are more likely to spill across caste-group-specific product markets. We classify districts above the median isolation index as segregated and restrict column (9) to the unsegregated half. The positive and significant LC coefficient in the unsegregated sample is consistent with demand-side spillovers operating more strongly where caste groups share local markets.

A.6.5.5 Firm characteristics. Column (12) restricts to large firms, defined as those with base-year (2005) gross output above the sample median. Larger firms may be better positioned to capture demand from other caste groups, and indeed the LC coefficient is the largest in the table. Column (13) restricts to 4-digit NIC industries where the share of LC-owned firms is below the sample median. In sectors not dominated by LC firms, LC producers are more likely to serve cross-caste demand rather than only LC consumers; the LC coefficient remains positive and significant.

A.6.5.6 Cross-column stability. Across all sample splits, the coefficients on Rainfall shock $\times$ LC and Rainfall shock $\times$ MC remain positive and significant. The p-values reported in the last two rows test the null hypothesis that each column's coefficient equals the benchmark in column (2); most of the differences are not statistically significant, confirming that the revenue response is stable across subsamples.

Table A.12: Firm-Level Revenue Response to Rainfall Shocks

	(1) NIC4 × Yr	(2) Prod × Yr	(3) Prod × Yr	(4) LCE	(5) Rest	(6) NT	(7) Low prod. price range	(8) Low prod. price SD	(9) Unseg. districts	(10) Low RM price range	(11) Low RM price SD	(12) Large firms	(13) Non-LC sectors
Rainfall shock × MC	0.099**	0.093***	0.046***	0.033*	0.048***	0.048***	0.031***	0.039***	0.056***	0.031***	0.034***	0.068***	0.058***
	(0.043)	(0.035)	(0.012)	(0.019)	(0.011)	(0.012)	(0.011)	(0.010)	(0.016)	(0.011)	(0.010)	(0.013)	(0.016)
Rainfall shock × LC	0.112**	0.111***	0.047***	0.018	0.051***	0.050***	0.036***	0.040***	0.056***	0.039***	0.038***	0.070***	0.047***
	(0.048)	(0.038)	(0.014)	(0.020)	(0.014)	(0.015)	(0.013)	(0.013)	(0.020)	(0.014)	(0.012)	(0.020)	(0.016)
p-value: MC = Col. 2	0.909	.	0.200	0.128	0.219	0.213	0.089	0.131	0.332	0.089	0.104	0.496	0.353
p-value: LC = Col. 2	0.983	.	0.115	0.029	0.140	0.137	0.062	0.076	0.203	0.073	0.069	0.345	0.118
Observations	1,208,106	1,205,165	1,115,322	258,686	856,624	439,012	813,326	826,180	549,140	796,426	823,114	551,339	545,558
R^2	0.457	0.560	0.833	0.813	0.845	0.847	0.828	0.844	0.814	0.828	0.845	0.712	0.812
Firm age			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm chars × Caste			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Revenue Quartile FE			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caste FE	✓	✓											
Dist × Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product × Year FE		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NIC4 × Year FE	✓												

Notes: The dependent variable is log gross output. The sample covers MSME firms from 2005–2007. HC (upper castes) is the omitted group; MC = OBC; LC = SC/ST. Column (1) uses NIC4 × year FE with caste FE; column (2) uses product × year FE with caste FE. Columns (3)–(13) add firm-level controls: firm age, base-year (2005) revenue quartile FE, and firm characteristics (power source, organization type, nature of operation, management type, accounting status, technology know-how) interacted with caste. Columns (4)–(13) restrict to subsamples as labeled (see text). All columns include district × year FE, analytic weights, and district-clustered standard errors. The p-value rows test H_0: coefficient in column j equals the benchmark in column (2). *, **, *** denote significance at the 10%, 5%, and 1% levels.

A.6.6 Fact 4.3: Workforce Composition in Competitive Markets

We examine whether HC-owned firms adjust their workforce composition in response to local demand conditions, and whether this adjustment depends on the degree of caste-based competition in the firm's product market. Using Equation (5), we regress the LC employee share of HC-owned firms on the interaction of the rainfall shock with an indicator for whether the firm operates in a competitive market—defined as a district $\times$ product (or district $\times$ NIC-4 industry) market where the HC share of activity falls below a given threshold.

Figure A.4 reports the interaction coefficient θ across alternative definitions of competitiveness. A market is classified as competitive if the HC share of firms, revenues, or employment is below x%, for $x \in \{50, 60, 70, 80\}$. Panel (a) defines markets at the district $\times$ product code level; Panel (b) uses the district $\times$ NIC-4 industry level. Across all definitions, the interaction coefficient is positive, indicating that HC firms in competitive markets hire relatively more LC workers when local LC demand rises. The effect is precisely estimated at the 50–70% thresholds and becomes noisier at 80%, as the "competitive" group expands to include markets with substantial HC presence.

These results support the demand channel: in markets where HC firms face competition from LC and MC producers, they respond to positive LC demand shocks by diversifying their workforce toward LC employees. The absence of such a response in HC-dominated markets—where cross-caste competition is limited—argues against a pure labour supply interpretation, since rainfall-induced changes in LC labour supply operate at the district level and do not vary with product-market concentration.

(a) Product Code (b) NIC-4 Industry

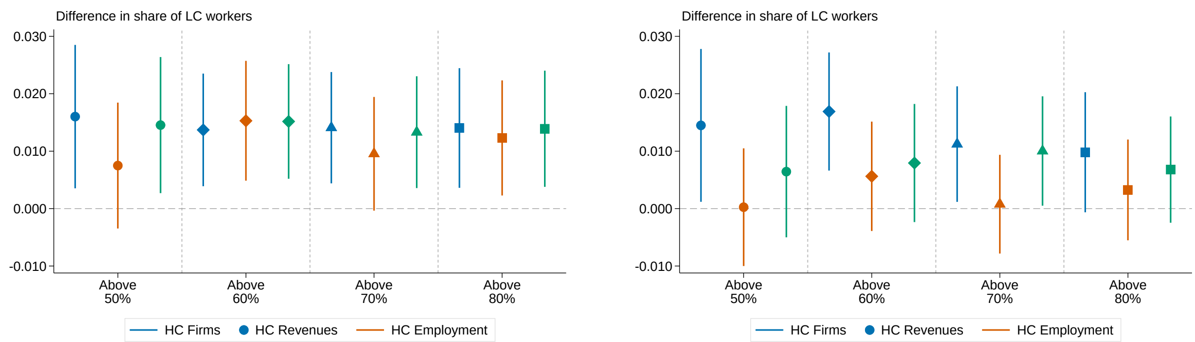


Figure A.4: LC Employee Share Among HC-owned Firms in Competitive Markets

Notes. The figure presents the coefficient θ from regressions using Equation (5) among HC-owned firms. The dependent variable is the share of LC (SC/ST) employees in total firm employment. A market is classified as competitive if the HC share of activity falls below x percent, where $x \in \{50, 60, 70, 80\}$. HC market share is measured by the number of firms (blue circles), revenues (orange circles), or employment (green circles). Marker shapes vary by threshold: circles (50%), diamonds (60%), triangles (70%), squares (80%). Panel (a) defines markets at the district $\times$ product code level; Panel (b) at the district $\times$ NIC-4 industry level. Bands represent 90% confidence intervals. All regressions include district, NIC-4, and caste fixed effects, with analytic weights and standard errors clustered at the district level.

Workforce composition in contact-intensive sectors. We also examine whether the demand channel operates more strongly in sectors where customer–employee interaction is prevalent. Table A.13 lists the NIC-2 industries included in each definition of contact-intensive sectors, along with the share of firms by caste. The baseline definition includes wood products, furniture manufacturing, construction, motor vehicle trade, wholesale and retail trade, hotels and restaurants, travel agencies, post and telecommunications, and computer services. The broad definition additionally includes real estate, machinery rental, education, and health services, as well as firms classified as service-oriented by activity nature.

Table A.13: Contact-Intensive Sector Definitions and Firm Composition

		Definition		Share of firms (%)			
NIC-2	Industry	Baseline	Broad	All	HC	LC	MC
20	Wood products	✓		4.6	3.1	5.8	5.7
36	Furniture mfg.	✓	✓	7.1	4.7	9.4	6.9
45	Construction	✓	✓	0.0	0.0	0.0	0.0
50	Motor vehicle trade	✓	✓	2.6	2.6	2.9	1.8
51	Wholesale trade	✓	✓	0.0	0.1	0.0	0.0
52	Retail trade	✓	✓	9.2	8.0	8.4	15.8
55	Hotels & restaurants	✓	✓	0.1	0.1	0.1	0.1
63	Travel agencies	✓	✓	0.2	0.5	0.1	0.0
64	Post & telecom	✓	✓	0.8	1.1	0.6	0.4
70	Real estate		✓	0.0	0.0	0.0	0.0
71	Machinery rental		✓	0.0	0.0	0.0	0.0
72	Computer services	✓	✓	0.7	1.0	0.4	0.4
80	Education		✓	0.0	0.0	0.0	0.0
85	Health & social		✓	0.1	0.1	0.1	0.1
	Total (Baseline)			25.4	21.1	27.8	31.2
	Total (Broad)			20.9	18.1	22.1	25.6

Notes. The table lists NIC-2 industries included in each contact-intensive definition. *Baseline* corresponds to the definition used in the main text. *Broad* is an expanded definition that additionally includes real estate, machinery rental, education, health and social services, and firms with a service-oriented activity nature. Share of firms is computed using sampling multipliers. HC = upper castes; LC = SC/ST; MC = OBC.

Table A.14 reports robustness results using the broad definition of contact-intensive sectors. The interaction coefficient remains positive and significant under both definitions: 0.010 ($p < 0.05$) under the baseline and 0.011 ($p < 0.10$) under the broad definition in the full sample. In the HCE subsample, the effects are larger (0.013 and 0.014, both $p < 0.05$).

Table A.14: LC Employee Share: Robustness to Contact-Intensive Definition

	Dep. var.: Share of LC workers			
	Baseline		Broad	
	(1) All	(2) HCE	(3) All	(4) HCE
Rshock × Contact	0.010**	0.013**		
	(0.005)	(0.005)		
Rshock × Contact			0.011*	0.014**
			(0.006)	(0.006)
District FE	✓	✓	✓	✓
NIC-4 FE	✓	✓	✓	✓
Observations	179,958	130,953	179,958	130,953
R^2	0.182	0.203	0.183	0.203
Mean LC share	0.096	0.095	0.096	0.095

Notes. The dependent variable is the share of LC (SC/ST) employees in total firm employment among HC-owned firms. *Rshock* is the 2006 rainfall shock. *Contact* is an indicator for contact-intensive industries. Columns (1)–(2) use the baseline definition; Columns (3)–(4) use the broad definition (see Table A.13). *HCE* denotes high consumption elasticity sectors. All regressions include district and NIC-4 industry fixed effects. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.15: LC Employee Share: Robustness to Firm-Level Controls

	Dep. var.: Share of LC workers							
	Competition		Competition + Controls		Contact		Contact + Controls	
	(1) All	(2) HCE	(3) All	(4) HCE	(5) All	(6) HCE	(7) All	(8) HCE
Competition	-0.024***	-0.015***	-0.012***	-0.006*				
	(0.004)	(0.004)	(0.004)	(0.004)				
Rshock × Competition	0.016**	0.016**	0.015**	0.014**				
	(0.008)	(0.008)	(0.007)	(0.007)				
Rshock × Contact					0.010**	0.013**	0.006	0.009**
					(0.005)	(0.005)	(0.004)	(0.004)
Firm controls			✓	✓			✓	✓
District FE	✓	✓	✓	✓	✓	✓	✓	✓
NIC-4 FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	179,958	130,953	170,519	123,084	179,958	130,953	170,519	123,084
R^2	0.184	0.203	0.197	0.210	0.182	0.203	0.196	0.210

Notes. The dependent variable is the share of LC (SC/ST) employees in total firm employment among HC-owned firms. *Competition* is an indicator equal to one if the HC share of firms in the district × product code market is below 50 percent. *Contact* is an indicator for contact-intensive industries (baseline definition). *HCE* denotes high consumption elasticity sectors. Columns (1)–(2) and (5)–(6) replicate the baseline specifications without firm controls. Columns (3)–(4) and (7)–(8) add firm-level controls: firm age, power source, organization type, nature of operation, management type, accounting status, and technology know-how. All regressions include district and NIC-4 industry fixed effects. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1.

A.7 Foreign Demand Shocks

The firm survey provides data on the value of exports in the years 2006 and 2007. 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 product codes. We manually developed a cross-walk between MSME product codes and HS product codes. We focus 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.[13]

Consider an exporter that produces a product p at time t. We observe over the entire period product p 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 p into destination d from all countries except India at time t. Thus, X_{dpt}^{-I} 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:

$$FD_{pt} = \sum_{d \in \Xi_d} \overline{s_{dp}} \cdot \Delta \log X_{dpt}^{-I}. \tag{A.1}$$

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 $\overline{s_{dp}} \equiv \frac{X_{dp}^{I}}{\sum_{d'} X_{d'p}^{I}}$ 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 situations where trade flows switch from zero to a positive number (a common feature of international trade

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

data).

Table A.16 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% and revenues by 0.58%. The cross-sectional results are also qualitatively similar – revenues and employment respond positively to a foreign demand shock. However, the coefficients on 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 A.16: 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
FD_{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^2	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 product level, *** p<0.01, ** p<0.05, * p<0.1.

B Model Derivations

B.1 Derivation of the Export Cutoff Ratio

From the zero-profit conditions (13) applied to the domestic market ($s' = s$) and an export market ($s' \neq s$) for a group-s firm:

$$\left(\frac{w_s}{z^*_{d,s} \cdot P_{D,s}}\right)^{1-\sigma} \cdot \frac{(1-a)(I_s - \bar{h})}{\sigma} \cdot \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} = f_d \cdot w_s, \tag{B.2}$$

$$e^{-\beta_{s',s}} \left(\frac{(1+g_{k,s,s'})\, w_{s'}}{z^*_{x,s,s'} \cdot P_{D,s'}}\right)^{1-\sigma} \cdot \frac{(1-a)(I_{s'} - \bar{h})}{\sigma} \cdot \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} = c_{x,s,s'} \cdot w_{s'}. \tag{B.3}$$

Dividing (B.3) by (B.2) and using $w_s = w_{s'} = 1$:

$$e^{-\beta_{s',s}} (1+g_{k,s,s'})^{1-\sigma} \left(\frac{z^*_{d,s}}{z^*_{x,s,s'}}\right)^{1-\sigma} \left(\frac{P_{D,s}}{P_{D,s'}}\right)^{1-\sigma} \frac{I_{s'} - \bar{h}}{I_s - \bar{h}} = \frac{c_{x,s,s'}}{f_d}. \tag{B.4}$$

Since $\sigma > 1$, rearranging and using the demand index $\kappa_s = (1-a)(I_s - \bar{h})\, P_{D,s}^{\sigma-1}$:

$$z^*_{x,s,s'} = z^*_{d,s} \cdot (1+g_{k,s,s'}) \cdot \left(\frac{c_{x,s,s'}}{f_d}\right)^{\frac{1}{\sigma-1}} \cdot e^{\frac{\beta_{s',s}}{\sigma-1}} \cdot \left(\frac{\kappa_s}{\kappa_{s'}}\right)^{\frac{1}{\sigma-1}}. \tag{B.5}$$

B.2 Productivity Distribution and Aggregation

Firms draw z from a bounded Pareto distribution with CDF

$$G(z) = \frac{1-(z_{\min}/z)^{\eta}}{1-(z_{\min}/z_{\max})^{\eta}}, \qquad z \in [z_{\min}, z_{\max}], \tag{B.6}$$

where $\eta > 0$ is the shape parameter and $z_{\max} < \infty$ is the upper bound. The finite upper bound ensures that all moments $\mathbb{E}[z^k]$ are finite for any k, so the standard unbounded-Pareto restriction $\eta > \sigma - 1$ need not be imposed. In our calibration, $\eta = 3.1$ and $\sigma = 5$, giving $\eta/(\sigma-1) = 0.775$.

Define:

$$\Phi_{s',s} \equiv \frac{1}{1-G(z^*_{d,s'})} \int_{z^*_{s',s}}^{z_{\max}} z^{\sigma-1}\, dG(z), \tag{B.7}$$

where $z^*_{s,s} = z^*_{d,s}$ for domestic firms and $z^*_{s',s} = z^*_{x,s',s}$ for exporters ($s' \neq s$). For domestic firms, $\Phi_{s,s}$ is the conditional expectation $\mathbb{E}[z^{\sigma-1} \mid z \geq z^*_{d,s}]$. For exporters, $\Phi_{s',s}$ equals the conditional probability of exporting times the conditional expectation of $z^{\sigma-1}$ among exporters: $\Phi_{s',s} = \Pr(z \geq z^*_{x,s',s} \mid z \geq z^*_{d,s'}) \cdot \mathbb{E}[z^{\sigma-1} \mid z \geq z^*_{x,s',s}]$.

When $\eta > \sigma - 1$, the unnormalised integral admits the closed form $\int_{z^*}^{z_{\max}} z^{\sigma-1}\, dG(z) = \frac{\eta}{\eta-(\sigma-1)}\, z_{\min}^{\eta}\, (z^*)^{-[\eta-(\sigma-1)]}$. When $\eta \le \sigma - 1$, the integrals are computed numerically on the bounded support $[z_{\min}, z_{\max}]$.

B.3 Solution Algorithm

Groups differ in endowments (L_s), and all bilateral parameters ($\beta_{s,s'}$, $c_{x,s,s'}$, $g_{k,s,s'}$) are pair-specific. Normalise $w_s = P_H = 1$ for all s, which holds as long as all groups produce the homogeneous good. The equilibrium is solved by iterating over the demand indices $\{\kappa_s\}_{s=1}^{3}$:

1. **Initialise** demand indices $\{\kappa_s\}_{s=1}^{3}$.
2. **Cutoff ratios**: Given $\{\kappa_s\}$, compute $z^*_{x,s,s'}/z^*_{d,s}$ pair-by-pair from equation (14).
3. **Free entry** $\Rightarrow \{z^*_{d,s}\}$
4. **Bilateral** Φ: Compute $\{\Phi_{s',s}\}$ from equation (B.7).
5. **Goods market clearing and income**: Solve the D-sector goods market clearing condition (19) for each market s, jointly with the profit identities $I_s = w_s L_s + \Pi_s$, for $\{M_s, I_s\}_{s=1}^{3}$. Labour market clearing determines H_s^{prod} residually.
6. **Price indices**: Compute $\{P_{D,s}\}$ from equation (17).
7. **Update**: Recompute $\kappa_s = (1-a)(I_s - \bar{h})\, P_{D,s}^{\sigma-1}$. If converged, stop; otherwise return to step 2.

Upon convergence, compute real income using equation (21).

B.4 D-Sector Productivity and Real Income

D-sector real output in market s is $Q_{D,s} = (1-a)(I_s - \bar{h})/P_{D,s}$. The D-sector productivity measure used in Figure 6(b) is real differentiated-good output per unit of D-sector labour:

$$\text{D-sector productivity} = \frac{\sum_s Q_{D,s}}{\sum_s L_{D,s}}, \tag{B.8}$$

where $L_{D,s} = L_s - H_s^{\text{prod}}$ is the labour allocated to the differentiated-good sector in market s. Economy-wide real income uses the non-homothetic cost-of-living index:

$$\text{Real income p.c.} = \frac{\sum_s (I_s - \bar{h})/P_{D,s}^{1-a}}{\sum_s L_s}. \tag{B.9}$$

C Quantitative Analysis

C.1 Revenue Elasticity Decomposition

We derive the partial-equilibrium decomposition used to identify β. A firm from group s with productivity z selling in market s' earns revenue:

$$r(z,s,s') = \Psi_{s',s} \left(\frac{\sigma}{\sigma-1} \cdot \frac{\mathcal{C}_{s,s'}}{z} \right)^{1-\sigma} P_{D,s'}^{\sigma-1} E_{D,s'}, \tag{C.10}$$

where $E_{D,s'} = (1-a)(I_{s'} - \bar{h})$ is group-s' differentiated-good expenditure. Revenue in each market is linear in $E_{D,s'}$.

In the partial-equilibrium exercise, we shock group incomes ($I_{s'} \to I_{s'}^{PE}$) while holding firm selection (the set of active firms and their export status) and price indices $P_{D,s'}$ fixed. Since $r(z,s,s')$ is proportional to $E_{D,s'}$, the PE revenue in market s' scales as:

$$r^{PE}(z,s,s') = r(z,s,s') \cdot \frac{E_{D,s'}^{PE}}{E_{D,s'}}. \tag{C.11}$$

Summing across destination markets, the total PE revenue of a firm (z,s) is:

$$R^{PE}(z,s) = \sum_{s'} r^{PE}(z,s,s') = \sum_{s'} r(z,s,s') \cdot \frac{E_{D,s'}^{PE}}{E_{D,s'}}. \tag{C.12}$$

The average revenue growth for group-s firms is:

$$\bar{R}_s^{PE} = \sum_z R^{PE}(z,s)\, f(z|s) = \sum_{s'} \frac{E_{D,s'}^{PE}}{E_{D,s'}} \sum_z r(z,s,s')\, f(z|s) = \sum_{s'} \frac{E_{D,s'}^{PE}}{E_{D,s'}} \cdot \bar{R}_{s,s'},$$

where $\bar{R}_{s,s'} = \sum_z r(z,s,s') f(z|s)$ is the average revenue from market s' and $f(z|s)$ is the productivity distribution among active (surviving) group-s firms. Dividing both sides by $\bar{R}_s = \sum_{s'} \bar{R}_{s,s'}$:

$$\frac{\bar{R}_s^{PE}}{\bar{R}_s} = \sum_{s'} \bar{\omega}_{s,s'} \cdot \frac{E_{D,s'}^{PE}}{E_{D,s'}},$$

where $\bar{\omega}_{s,s'} = \bar{R}_{s,s'}/\bar{R}_s$ is the revenue share from market s'. This decomposition is exact—not a first-order approximation—because PE revenue is linear in expenditure.

The relative revenue growth (LC vs HC) is therefore:

$$\log \frac{\bar{R}_{LC}^{PE}}{\bar{R}_{LC}} - \log \frac{\bar{R}_{HC}^{PE}}{\bar{R}_{HC}} = \log\left(\sum_{s'} \bar{\omega}_{LC,s'} \cdot \frac{E_{D,s'}^{PE}}{E_{D,s'}} \right) - \log\left(\sum_{s'} \bar{\omega}_{HC,s'} \cdot \frac{E_{D,s'}^{PE}}{E_{D,s'}} \right).$$

Since only LC and MC incomes are shocked ($E^{PE}_{D,HC} = E_{D,HC}$), the relative elasticity depends on $(\bar{\omega}_{LC,LC} - \bar{\omega}_{HC,LC})$ and $(\bar{\omega}_{LC,MC} - \bar{\omega}_{HC,MC})$. Higher β increases $\bar{\omega}_{LC,LC}$ (LC firms earn more from their own market) and decreases $\bar{\omega}_{HC,LC}$ (HC firms earn less from the LC market), amplifying the relative response. This produces the monotonic relationship between β and the relative revenue elasticity shown in Figure 5.

C.2 Robustness: Alternative $(a, \bar{h})$ Calibrations

The baseline calibration sets the homogeneous-good expenditure share $a = 0.30$ and subsistence requirement $\bar{h} = 0.30$. These two parameters jointly control the strength of non-homothetic demand amplification: lower a and higher $\bar{h}$ increase the income elasticity of differentiated-good demand, particularly for poorer groups. We re-estimate 13 of the 15 internally calibrated parameters—β, six g_k, and six c_x, holding η and f_d fixed at their baseline values—for four alternative $(a, \bar{h})$ combinations, using the same Sobol-based search algorithm and moment targets. Across all specifications, the counterfactual gains from removing identity frictions are similar (aggregate real income increases by 1.050–1.096×). These results confirm that the baseline findings are not driven by the particular choice of $(a, \bar{h})$, see Table C.17.

C.3 Extension: Iceberg Transportation Costs

The baseline model assumes that all frictions in cross-group trade operate through demand bias ($\beta_{s,s'}$), hiring wedges ($g_{k,s,s'}$), and fixed export costs ($c_{x,s,s'}$). In this extension, we introduce a fourth channel: standard iceberg transportation costs that capture the physical or logistical cost of delivering goods across caste-segmented markets.

C.3.1 Setup

Let $\tau_{s,s'} \geq 1$ denote the iceberg cost for a group-s firm to deliver goods to market s', with $\tau_{s,s} = 1$ (no cost within the home market). To deliver one unit to consumers in market s', a firm from group s must ship $\tau_{s,s'}$ units. The $\tau_{s,s'} - 1$ units melt in transit and represent a real resource cost.

The unit cost of serving market s' becomes:

$$\mathcal{C}_{s,s'} = \begin{cases} w_{s'} & \text{if } s = s', \\ \tau_{s,s'} \cdot (1 + g_{k,s,s'})\, w_{s'} & \text{if } s \neq s'. \end{cases} \tag{C.13}$$

CES markup pricing gives:

$$p(z, s, s') = \frac{\sigma}{\sigma - 1} \cdot \frac{\tau_{s,s'} \cdot (1 + g_{k,s,s'}) \cdot w_{s'}}{z}, \qquad s \neq s'. \tag{C.14}$$

Revenue of a group-s firm with productivity z selling in market s':

$$r(z, s, s') = \Psi_{s',s} \left(\frac{\sigma}{\sigma - 1} \cdot \frac{\tau_{s,s'} \, (1 + g_{k,s,s'}) \, w_{s'}}{z \cdot P_{D,s'}} \right)^{1-\sigma} (1 - a)(I_{s'} - \bar{h}). \tag{C.15}$$

Dividing the export zero-profit condition by the domestic one yields the bilateral export cutoff:

$$z^*_{x,s,s'} = z^*_{d,s} \cdot \tau_{s,s'} \cdot (1 + g_{k,s,s'}) \cdot \left(\frac{c_{x,s,s'}}{f_d} \right)^{\frac{1}{\sigma-1}} \cdot e^{\frac{\beta_{s',s}}{\sigma-1}} \cdot \left(\frac{\kappa_s}{\kappa_{s'}} \right)^{\frac{1}{\sigma-1}}. \tag{C.16}$$

Compared with equation (14), the iceberg cost $\tau_{s,s'}$ enters multiplicatively and raises the export cutoff. The four frictions now compound: taste bias, hiring wedges, fixed costs, and transportation costs each independently raise $z^*_{x,s,s'}$, so that only the most productive firms overcome all four barriers to serve out-group markets.

The CES price index in market s becomes:

$$P_{D,s}^{1-\sigma} = \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \left[M_s \, \Phi_{s,s} + \sum_{s' \neq s} M_{s'} \, e^{-\beta_{s,s'}} \, (\tau_{s',s} \, (1 + g_{k,s',s}))^{1-\sigma} \, \Phi_{s',s} \right]. \tag{C.17}$$

Since $\sigma > 1$, the term $\tau_{s',s}^{1-\sigma} < 1$ reduces the contribution of imported varieties to the price index, making the market effectively more closed.

C.4 Composite Caste Resistance

With the iceberg cost, the caste-resistance term from equation (15) generalises to:

$$T_{s,s'} = e^{\beta_{s',s}/(\sigma-1)} \cdot \tau_{s,s'} \cdot (1 + g_{k,s,s'}), \qquad T_{s,s} = 1. \tag{C.18}$$

Each component enters multiplicatively:

- $e^{\beta_{s',s}/(\sigma-1)}$: demand-side identity barrier, converted to cost-equivalent;
- $\tau_{s,s'}$: physical transportation and logistics cost;
- $(1 + g_{k,s,s'})$: labour-market hiring wedge (deadweight loss).

The export cutoff (C.16) and price index (C.17) take the same compact form as in the baseline—equations (16) and (17)—with $T_{s,s'}$ now incorporating all three channels.

C.5 Discussion

The hiring wedge captures frictions specific to managing a cross-caste workforce (network frictions, information asymmetries), while $\tau_{s,s'}$ captures costs of reaching consumers across socially segmented markets (distribution networks, trust-based supply chains, geographic co-location of caste communities). When $\tau_{s,s'} = 1$ for all pairs, the model reduces to the baseline specification in Section 5.

In the baseline calibration, we set $\tau_{s,s'} = 1$ and absorb all variable trade frictions into $g_{k,s,s'}$. With additional data on distribution costs or intermediary margins across caste-segmented markets, one could separately identify $\tau_{s,s'}$ from $g_{k,s,s'}$—for instance, by exploiting variation across products that differ in transportability but share similar hiring patterns.

C.5.1 Empirical Estimates of Transportation Costs in India

Asturias et al. (2019) estimate bilateral iceberg costs between Indian states using price differences of goods produced by national monopolists. The authors fit a smoothed monotonic cubic to decile coefficients: $g(x) = 0.9 + 0.176\,x - 0.0317\,x^2 + 0.002\,x^3$, where $x \in \{1, \ldots, 10\}$ denotes the decile of effective distance. Normalising the first decile to unity, implied iceberg costs range from $\tau = 1.00$ to 1.42 (0–42% ad valorem) across distance deciles. Van Leemput (2021) estimates internal trade barriers account for 13–70% of total trade costs across states.

C.5.2 Implications for our model.

Both studies estimate transportation costs between geographically separated regions (states). In our model, caste groups are not geographically separated but co-located within the same local economy, so the physical transportation component of $\tau_{s,s'}$ is likely small. We therefore consider two modest values: $\tau = 1.03$ (3% ad valorem), which corresponds to the estimated cost between neighbouring states such as Delhi and Haryana in Asturias et al. (2019); and $\tau = 1.10$ (10% ad valorem), which falls between the 2nd and 3rd distance deciles of the smoothed interstate estimates above. We recalibrate the augmented model at each value. In the baseline ($\tau = 1$), removing all identity-based frictions ($\beta = 0$, $g_k = 0$) raises aggregate real income to 1.076 times the benchmark. With $\tau = 1.03$, this ratio falls to 1.061; with $\tau = 1.10$, it falls to 1.039. The reduction occurs because the iceberg cost τ remains even after identity barriers are removed, limiting the gains from eliminating β and g_k alone, see Table C.17.

Table C.17: Robustness of Calibrated Parameters

	$(a, \bar{h})$ combination				Iceberg cost τ	
	$(0.20, 0.20)$	$(0.20, 0.40)$	$(0.40, 0.20)$	$(0.40, 0.40)$	$\tau = 1.03$	$\tau = 1.10$
No taste bias, no hiring wedges ($\beta = 0$, $g_k = 0$), rel. to benchmark						
LC	1.133	1.127	1.098	1.072	1.091	1.043
MC	1.125	1.126	1.092	1.094	1.087	1.043
HC	1.054	1.016	1.040	1.011	1.028	1.034
Aggregate	1.096	1.073	1.071	1.050	1.061	1.039
Var(log r.i. p.c.), benchmark	0.0067	0.0346	0.0053	0.0278	0.0150	0.0141
Var(log r.i. p.c.), no friction	0.0026	0.0200	0.0024	0.0197	0.0093	0.0132

Notes. Columns 1–4 re-estimate 13 of the 15 internally calibrated parameters (β, six g_k, six c_x; η and f_d held fixed) for different $(a, \bar{h})$ combinations (baseline: $a = 0.30$, $\bar{h} = 0.30$). Columns 5–6 augment the baseline model with iceberg costs τ on cross-caste trade. Each column reports real income per capita under the counterfactual with $\beta = 0$ and $g_k = 0$, relative to the calibrated benchmark. The last two rows report the population-weighted variance of log real income per capita under the benchmark and no-friction counterfactual, respectively.