Social Norms, Product Demand and Firm Size*

Sampreet S. Goraya† Akhil Ilango‡

This draft: September 13, 2023

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

Abstract

This paper examines the impact of caste norms on product demand, firm size and real income in rural India. First, we establish changes in regional rainfall intensity as an asymmetric demand-shifter to members of different castes. Specifically, when consumers from historically disadvantaged low-ranked communities in the caste system (LC) experience a positive shift in income due to good rainfall, relative to others, they increase their spending across various non-agricultural product categories, whose supply is not directly affected by rainfall. Second, we use novel data on the caste of firm owners and the caste composition of employees within a firm to document that the increase in LC households' demand leads to higher growth for the firms owned by members of the same caste category, relative to others. Motivated by these empirical findings, we develop a theoretical framework where firms sell products across castes, and consumers' taste for products depends on the quality and the caste of the producer. And in response, firms invest to influence the consumer's taste by hiring workers from the target consumers' caste. We plan to quantify the role of caste-specificity in demand in lowering cross-caste sales, lowering firm size, raising the proportion of small firms, and lowering real income.

Keywords: Caste norms, Trade, Product quality, Firm size.
JEL Codes: O11, D43, D61.

*We thank Rodrigo Carril, Antonio Ciccone, Rosa Ferrer, Sylvie Lambert and Sandro Shelegia as well as the participants at the CREI international lunch, Paris School of Economics development seminar, Stockholm School of Economics, Mannheim applied seminar for helpful discussions and comments.

†Department of Economics, Stockholm School of Economics. Email: sampreet.goraya@hhs.se

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

1 Introduction

A large number of firms are small and grow slowly in developing countries (Tybout, 2000; Hsieh and Klenow, 2014). A number of supply-side explanations have been put forward such as credit constraints, labor market regulations, taxes and subsidies, among others. In this paper, we focus on demand-side determinants of firm size. In particular, we highlight how social norms can influence the demand for certain products relative to others. Such demand asymmetries can restrict the scale of firms and favor the survival of small firms that produce low-quality goods.

A natural starting point for such analysis is a setting where social norms can play a crucial role for economic outcomes. Hence, we turn our attention to the caste system in India. Caste is inherited at birth and determines one's social identity. Historically, the caste system restricted inter-caste interactions and promoted discriminatory practices towards low-ranked communities (LC). Despite immense socioeconomic changes in the last few decades, caste remains a salient feature of Indian society. Previous work has argued that consumers' identity shapes their preference for goods and services (Atkin, Colson-Sihra, and Shayo, 2021). We build on this literature and argue that adherence to caste norms translates into a strong preference for goods that are produced by firm owners of the same caste. This is particularly relevant for firms in India as the product demand tends to be local.[1] Therefore, the boundary of a firm owned by a member of a certain caste is limited to the size of that caste within that region.

In order to highlight this channel, we exploit data from the Micro, Small and Medium Enterprises (MSME) survey of 2006-2007. This dataset contains a representative sample of MSMEs together with an exhaustive list of balance sheet variables, product prices and quantities, and information on the caste of the enterprise owner and employees, a feature missing in other commonly used firm-level datasets in India (see Goraya, 2023). We show that the LC consumers have lower income and consume less, and the firms owned by LC are relatively small. The average firm size is increasing in the consumption share of LC within a region. Looking at the employer-employee linkages, we find that over 75-80 percent of the employees in a firm belong to the same caste as the employer. The large firms have a relatively more heterogeneous caste composition of employees.

In our second string of evidence, we show that when there is an exogenous positive shift in income specifically to LC households, they respond by increasing their expenditure on manufacturing goods and services. We observe this increase across various product categories and industries. However, this increase in demand does not translate into sales uniformly, across firms owned by members of different castes. The LC firms exhibit a substantial increase in revenue, relative to HC firms. These evidences show that demand is

[1] See Jensen and Miller (2018).

segmented and that caste-linkages are an important driver of firms' demand and hence, firm size, in rural India. Further, we find that this increase in firm revenues is not driven by a change in prices or an increase in access to credit.

In order to rationalize these findings, we build a heterogeneous firm general equilibrium model that explains the decision of firms to sell to different castes and the caste composition of employees for small and large firms. Castes are heterogeneous in their size. HC are large (rich) and LC are relatively small (poor). Firms are heterogeneous in the quality of the product and the caste of the owner. Both of these attributes determine the product demand. Conditional on the quality, consumer dislikes the goods produced by ethnically distant castes relative to their own caste. While accessing consumers of different castes, firms incur a fixed cost. Firms that sell to many castes are positively selected and tend to be larger. However, the share of revenues from each caste depends on the caste of the owner. LC firms have a higher share of revenues from LC consumers relative to the HC firms.

Firms that indulge in cross-caste trade can also invest in reducing the distaste for their product by hiring employees of the target caste. However, becoming ethnically closer to one caste may come at the cost of becoming ethnically farther away from the other. Therefore, the firm's decision to enter a caste market not only depends on the income of the target caste but also on the distribution of income of all relevant castes. We show that, under certain conditions, the optimal ethnic location for the firm is a weighted average of the ex-ante ethnic distance, where weights are proportional to the income share.

The large firms' ability to influence consumer tastes dampens the effect of caste norms on trade and firm size. Large firms from LC, that also produce high-quality goods, have the incentive to be closer to richer castes. The ability to change consumer taste comes at the expense of hiring employees of the target castes. The cost is allowed to be heterogeneous across owners. If we assume that there is some discrimination in the labor market such that LC owners can not hire HC workers whereas HC owners can hire LC workers easily then the caste differences in firm size are magnified. Thus, consumer tastes influence the caste composition of the employees within a firm.

To analyze the effect of caste on firms through product demand, we combine data from India on individual-level wages from the NSS Employment and Unemployment surveys, household-level consumption from the NSS Consumer Expenditure surveys, and firm-level revenue, material input, wages, prices, and loans from the MSME survey. In order to assess the causal effect of local demand on firm growth and inefficiencies in input allocation, we use changes in the local rainfall as a proxy for exogenous shifts in local income, especially for the LC households.

To establish rainfall as an income shifter, we first estimate the effect of higher rainfall on individual wages.[2] Higher rainfall increases agricultural productivity and therefore, wages

[2]See also Jayachandran (2006), Kaur (2019), Santangelo (2019), Gupta (2020) for recent work on the effects of rainfall on the Indian labour market.

for workers involved in the agricultural sector. We show that a 1% increase in rainfall induces an increase in LC workers' daily wages by 7%, which is 3.3% more than the increase for HC workers. Further, since agriculture is the predominant sector of employment for individuals from LC households (42% of LC workforce compared to 28% of HC workforce), this worker-level effect can translate into a substantial shift in demand from the LC community, relative to others. We investigate this next.

To analyse the effect of the shift in income on consumption patterns across castes, we estimate the effect of higher rainfall on households' Monthly Per-Capita Expenditure ($mpce$). A 1% increase in rainfall leads to an average increase in MPCE by 10% for LC consumers and no change in MPCE for HC consumers. The rise in consumption is also observed when focusing only on non-agricultural sectors including, manufacturing, services, or durables. Therefore, although higher rainfall generates a direct impact on the supply of inputs and production predominantly in the agricultural sector, it indirectly induces an increase in demand in other sectors where it leaves no major direct effect on the supply-side. Motivated by this, we next investigate the indirect effects that higher rainfall causes on firms' input in other sectors, due to a rise in their demand.

To document the effect of a shift in LC households' demand on firms' outcomes, and to test the existence of caste-linkages in LC firms' demand, we next estimate the effect of higher rainfall on firm-level outcomes. We find that corresponding to the observations about $mpce$, there is an increase in the revenue of LC firms by 17.2%, relative to HC firms, for a 1% increase in rainfall. We find that the firms increase their use of intermediate inputs.

We documented the presence of such ethnic linkages and the resulting segregation in the product market. However, this overall segregation may be a combination of two underlying channels: (a) homophilic preferences, where consumers have a preference to buy from firms belonging to ethnically similar groups, and (b) geographical distance, where consumers have a cost to access products that geographically farther. We show that our framework allows us to include both of these forces and derive the baseline results. On the empirical side, we use the village level population data from the SHRUG database (Asher and Novosad, 2019) to measure geographical segregation within each district. We calculate the standard deviation of LC population share within a district and we omit observations of highly geographically segregated districts, that is those districts that belong to the top quartile of our measure of segregation. This exercise shows similar results to our baseline estimates. This suggests that geographical segregation may not be the driving factor behind the homophilic demand patterns.

Finally, we provide evidence to rule out alternative mechanisms that may lead to the same economic outcomes. We estimate the change in input prices due to higher rainfall and observe no significant effect. Considering input prices as a proxy for product quality, this result reassures us that changes in product quality are unlikely to be driving the demand effects. Further, no change in prices also shows that the rise in revenue is driven by a rise in quantity sold or growth in the size of the firms. In addition, to address concerns on the supply-side caste linkages (e.g., through loans from caste networks), we test the change in informal loans taken by firms and find no evidence of a significant change. Together, these observations indicate that it is mainly the channel of demand-side constraints that drives the growth patterns of LC firms that we observe in rural India.

1.0.0.1 Literature Review : This paper contributes to three broad strands of the literature, namely those of social norms and trade, of social norms and firm-level distortions, and of the demand-side determinants of firm growth.

First, this paper contributes to the literature on social norms and trade. Anderson (2011) shows that caste norms restrict trade across communities, Fujiy, Khanna, and Toma (2022) and Boken, Gadenne, Nandi, and Santamaria (2022) show that firms of similar castes are more likely to trade with each other. Desmet and Gomes (2023) shows how ethnic remoteness hampers market access. Our paper focuses on consumer-firm linkages and highlights how demand-side distortions may spillover to labor markets. We show that restricted market access across castes lowers firm size and real income.

Second, this paper contributes to the literature that links social norms and economic growth. Hsieh, Hurst, Jones, and Klenow (2019) and Cassan, Keniston, and Kleineberg (2021) study the distortions generated by race in the allocation of talent across space and occupations. Goraya (2023) study the macroeconomic implications of caste-based capital market distortions.[3] We focus on the macroeconomic implications of the product market distortion. We show that the product market distortions shape the labor allocation across firms.

Third, this paper adds to the growing body of evidence on the importance of demand-side factors for firm growth. Foster, Haltiwanger, and Syverson (2016) shows that building a consumer base is key for the growth of a firm, even if it is as efficient as other firms in terms of production. Hottman, Redding, and Weinstein (2016) uses bar-code data to show that "firm appeal" explains a lot of firm size heterogeneity in the US. Einav, Klenow, Levin, and Murciano-Goroff (2021) uses customer transaction data to show that customer acquisition accounts for a large portion of sales heterogeneity among firms in the US. Bernard, Dhyne, Magerman, Manova, and Moxnes (2022) shows that the number of final buyers explains a large majority of firm size heterogeneity in Belgium. While the above papers provide evidence on the importance of the extensive margin of demand for firm growth, our paper complements this literature by providing evidence from an emerging economy on the importance of the intensive margin of demand for firm growth, especially when the extensive margin faces resistance to change (due to caste in our setting). Our paper provides further

[3]There are several other contributions such as Banerjee and Munshi (2004) argue that capital is inefficiently allocated across communities in Tirupur in India. Munshi and Rosenzweig (2016) argue that caste norms dampen rural-urban migration. Oh (2021) study the interplay between identity and labor supply. Hjort (2014) shows that interethnic rivalries lower allocative efficiency.

evidence that restricted market access reduces firm scale and real income.

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

2 Institutional Background

The caste system in India is a form of social stratification that historically divided people into rigid hierarchical groups based on their occupation. For centuries, caste has been assigned to individuals at birth, irrespective of their parents' current occupation. It is propagated through endogamy, and has dictated customary social interaction and exclusionary practices.[4] In descending order of dominance, the hierarchy dictated by the system, based on historical occupation, is as follows: Brahmins (priests and teachers), Kshatriyas (rulers and soldiers), Vaishyas (merchants and traders), and the Sudras (labourers and artisans). Further, there are two additional groups that fall outside and below the caste system. The first one embodies the group of people traditionally known as Dalits.[5] The second group of people is known as the Scheduled Tribes.

Despite India's rapid economic development in recent decades, this social structure continues to function as a strong alternative to the underdeveloped market institutions. Therefore, to address the information and commitment problems, the Indian economy also is organised around castes.[6] Therefore, the groups lower in the hierarchy have also been subject to various forms of economic discrimination, including *barriers to access capital and firm creation*.

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

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

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

[6] See Munshi (2019) for an overview of the role of caste in the various facets of the Indian economy.

3 Theoretical Framework

In this section, we describe a model of within-district sales, in a district populated by members of different castes. There are $\mathcal{S}$ castes in the region and the ethnic distance between two castes denoted by s and s' is defined as Euclidean distance $d_{ss'}$ and by symmetry $d_{ss'} = d_{s's}$. We assume that $0 \leqslant d(s,s^{'}) \leqslant 1$, where $d_{ss'} = 0$ means zero ethnic distance and $d_{ss'} = 1$ means s and $s^{'}$ are ethnically most distant castes. We will define the ethnic distance more precisely in Section 3.3. We now describe the household sector and the production sector.

3.1 Households

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

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

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

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

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

where $C_{H,s}$ is the demand for homogeneous goods and $C_{D,s}$ is the demand for high-quality differentiated goods. $C_{D,s}$ is a bundle of differentiated varieties produced by firms of different castes at a cultural distance $d_{ss^{'}} \in [0,1]$. Further, we assume that households have CES preferences over endogenous Ω_s differentiated varieties, where σ denotes the elasticity of substitution between varieties. Each variety ω has two attributes, quality $z(\omega)$ and the caste of producer $s^{'}$. The $c(z(\omega),s,s^{'})$ denotes the units of consumption by household s of good quality $z(\omega)$ and produced by caste $s^{'}$, $q(z(\omega),s,s^{'})$ denotes the utility derived per unit good consumed, $p(z(\omega),s,s^{'})$ is the price. Further, it is assumed that a household needs a minimum amount of homogeneous goods $\overline{H}$ to survive. This implies that as households get richer their share of expenditure on homogeneous goods declines and expenditure on differentiated varieties increases. Household total income is denoted by I_s which is composed of total wage income and net profits from firms that belong to the same caste. The optimal expenditure and the demand for a variety with quality $z(\omega)$ is given by an iso-elastic demand curve:

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

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

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

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

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

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

3.2 Production Sector

3.2.1 Homogeneous good sector

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

3.2.2 Differentiated Good Sector

We assume that differentiated products are produced by a continuum of firms. Therefore, we use z as a marker of product quality. In this sector, firms produce with a Cobb-Douglas production technology with constant returns to scale, $y(z, s) = \mathcal{F}_s(\{\ell_e(s')\}_{s' \in S})$. The labour is the only input in production. The firms consider each ethnic market, say of caste s' at a distance $d_{ss'}$, as a separate market. In order to access this market, firms need to pay a fixed cost of $f^d_{ss'}$ every period.[7] Given consumers' demand, a firm with product quality z, and caste s solves the following profit maximization in each ethnic market at a cultural distance $d_{ss'}$.

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

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

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

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

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

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

3.2.3 Firm Scale

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

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

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

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

3.3 Endogeneous Ethnic Proximity

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

Let us define $\mathcal{X}_s = \{\mathcal{X}_{s,1}, ..., \mathcal{X}_{s,\mathcal{N}}\}$, a column vector, contains the Cartesian coordinates of caste s in the $\mathcal{N}$ dimensional space.[8] The euclidean distance between two castes represented by $\mathcal{X}_s$ and $\mathcal{X}_s'$ is $d_{ss'}^2 = \sum_{k=1}^{\mathcal{N}} (d_{ss',k})^2$, where $d_{ss',k} = \mathcal{X}_{s,k} - \mathcal{X}_{s',k}$ is the L1 distance in the k^{th} dimension. Further, let us define the distance moved by a firm owner that was born in a caste group s by a $\mathcal{N} \times 1$ column vector $\Delta\mathcal{X}_s = \{\Delta\mathcal{X}_{s,1}, ..., \Delta\mathcal{X}_{s,\mathcal{N}}\}$, where $\Delta d_s^2 = \sum_{k=1}^{\mathcal{N}} (\Delta\mathcal{X}_{s,k})^2$.[9] We assume firms pay a caste-dependent moving cost that is given by a function $\Phi(\Delta\mathcal{X}_s; \Gamma_s)$, where Γ_s is $\mathcal{N} \times 1$ column vector of parameters that disciplines the costs of moving. Here, we allow for the possibility that LC firms may pay higher or lower to move closer to HC consumers than HC firms pay to move closer to LC consumers.[10] Profits of the firm owner s that chooses to move is given by

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

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

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

where taste depends on the ex-post cultural distance between the firm and consumer $d_{ss',k}^* = d_{ss',k} - \Delta\mathcal{X}_{s,k}$.[11] The parameter $\hat{\beta}$ captures the taste elasticity with respect to caste distance ($d_{ss'}^2$). We use a first-order Taylor approximation of this function that gives us $\Psi(d_{ss'}, \Delta d_{s'})^\sigma \approx 1 - \beta \left(\sum_{k \in \mathcal{N}} (d_{ss',k} - \Delta\mathcal{X}_{s',k})^2 \right)$ with $\beta = \hat{\beta}\sigma$. This gives us a quadratic taste function that is reminiscent of Hotelling (1929) quadratic cost functions. The advantage of the quadratic taste function is that it gives us the closed-form solution to the optimal ethnic distance problem. Together, with this, we assume that the cost of moving in caste space is quadratic in

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

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

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

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

ethni distance and is given by

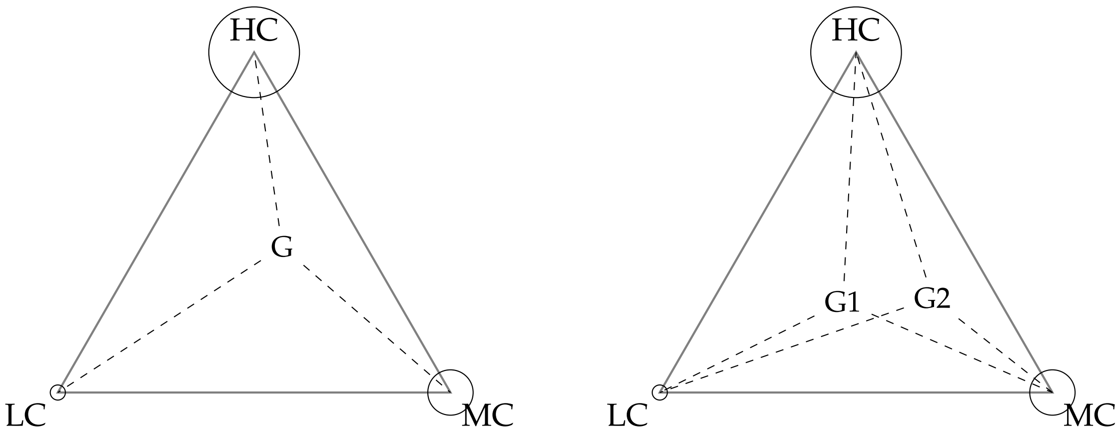
$$\Phi(\Delta\mathcal{X}_s;\Gamma_s) = \sum_{k\in\mathcal{N}} \gamma_{s,k}\Delta\mathcal{X}_{s,k}^2; \tag{13}$$

where $\gamma_{s,k}$ is the cost of moving along the dimension k for the firm owner of caste s.[12][13] We assume that these costs are paid in the wages of the skilled labor that belongs to the castes towards which the firm owner wants to move closer. Using Equation A.4, we can rewrite profits as

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

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

Figure 1: Optimal Ethnic Distance for Firms



(a) No Adjustment Costs

(b) Adjustment Costs

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

Proposition 1. *The optimal ethnic distance for a firm is a relative size-based weighted average of the*

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

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

distance between the caste of the firm owner and the caste of the consumer and is given by

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

Proof. See Proof in Appendix A. ■

3.4 Taste Shifters and Transportation Costs

In the previous section, we only focused on taste-specific demand shifters. However, this may be confounded by the transportation costs if consumers of different castes reside in different locations (segregated regions). Under such conditions, the trade across castes will be affected by both tastes and transportation costs.[14] Here, we derive the solution to the optimal caste identity under both of these distortions. We assume that physical distance (that captures transportation costs) and caste distance (that captures taste) are perfectly correlated. This is a reasonable assumption since LC communities tend to be segregated.[15] Therefore firm owners from LC communities may have a need to include transportation costs when selling to consumers of high castes. We follow the trade literature and assume transportation costs as an iceberg cost. We assume that there is no transportation cost to sell to your own caste (as firms and consumers are in the vicinity of each other). In this case, the price of a product, which depends on to whom a firm is selling, is given by

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

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

We assume that when firms change their ethnic identity, it also translates into a change in the physical distance and that they are perfectly correlated.[16] Further, we assume that transportation costs are quadratic in distance $\tau(d_{ss'}, \Delta d_s) = e^{\gamma\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s,k})^2\right)}$, where γ disciplines the rate at which the costs increase. The overall trade barrier for a firm can be summarized as composite $\Lambda(d_{ss'}, \Delta d_s) = \tau(d_{ss'}, \Delta d_s)^{1-\sigma}\Psi(d_{ss'}, \Delta d_s)^{\sigma}$. We use the first-order Taylor approximation of $\Lambda(d_{ss'}, \Delta d_{s'}) = 1 - \tilde{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k} - \Delta\mathcal{X}_{s',k})^2\right)$, where $\tilde{\beta} = \hat{\beta}\sigma + \gamma(\sigma - 1)$ that captures the strength of the cross-caste trade barriers. Under these assumptions, the solution is similar to Equation A.10, with the only difference being that the $\tilde{\beta}$, which captures the rate at which trade between two castes declines with ethnic and physical distance, is a

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

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

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

combination of taste parameter β and transportation cost parameter γ. We provide more details in Appendix A.[17]

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

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

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

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

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

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

Proof. See Proof in Appendix A. ■

In Proposition 2, we provide the expression for the revenue elasticity to a demand shock to caste s'. There are two effects, the direct effect and the indirect effect. The former comes from the increase in the size of the caste and thus the demand for the firm's product. As consumer gets richer, they demand more on differentiated varieties, thus the total expenditure on differentiated varieties increases.[18] Second, the indirect effect captures the fact that firms change their cultural proximity to caste s, when its income increases. It is easy to see that firm owners selling to own caste consumer, this effect is positive, that further increases the caste-specific revenues elasticity. Our framework nests the standard trade model, where micro trade elasticity is given by the standard term. Under fully exogenous cultural proximity with $d_{s's'} = 0$ and $d_{ss'} > 0$, the revenue elasticity to demand shocks is reduced to only the direct effect.

The elasticity of total firm-level revenue to an income shock is a product of caste-specific revenue elasticity and the share of revenue attributed to that specific caste. This implies that

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

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

firms that sell relatively more to the caste that experiences an income shock also witness a higher increase in overall revenues. Next, we are interested in how these elasticities respond to the changes in the trade resistance parameter $\tilde{\beta}$.

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

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

Proof. See Proof in Appendix A. ■

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

3.4.1 Firm Entry and Exit

To produce in the differentiated good sector, firms must pay a fixed entry cost, which is thereafter sunk. In equilibrium, with positive production of differentiated goods by firms of each caste, we require that the expected value of entry be equal to the sunk cost of entry f_e, which is paid in terms of skilled labor of that particular caste.[19] Firms draw their quality z from an exogenous distribution with CDF $G(z)$ and they enter the market if the profits are positive. If they do, they receive profits π every period they produce. Moreover, firms are risk neutral and face an exogenous probability of exit δ. The existing firms are replaced by the new entrants such that the firm distribution remains stationary. We denote by $\{L^e_s\}_{s\in\mathcal{S}}$ the mass of skilled workers used for entry and by $\{M^e_s\}_{s\in\mathcal{S}}$ the mass of potential entrants each period.

3.5 Equilibrium

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

1. Households maximize utility according to (1).

[19] We can allow for a more general structure of labor for entry costs. In our benchmark analysis, we try to capture the idea that castes that have low wages for skilled labor also tend to have lower costs of entry.

2. Producers of differentiated varieties maximize profits and charge the constant markup price given by Equation (8).

3. Product market clear for the homogeneous good and for each of differentiated goods.

4. The free entry condition holds.

5. Labor market clears for each caste.

4 Data and Measurement

We combine data from multiple sources to provide an empirical account of the effects of rainfall-induced changes in local income and the impact of caste linkages on the local economy. The strength of the compiled data set is that it contains micro-level data on the caste of both, firm owners and households, from across districts in India. It also contains information on the consumption behaviour, wages, demographics and firm production.

In this section, we introduce the data and provide a description of the economic conditions of individuals, households, and firms across different castes across districts in India.

4.1 Micro Small and Medium Enterprise (MSME) Data

The MSME data set consists of two parts: a census of registered MSMEs and a survey of unregistered MSMEs (recently used in Goraya, 2023).[20] In particular, the data set provides the geographical information, industry classification, balance sheet variables, and especially, the caste of the firm owner. There are two measures of capital stock in the data: the original value of investment in plant and machinery, and the market values of fixed assets. The total wage bill includes salaries and wages, allowances, and bonuses. We use *gross sales* as the measure of revenue or output. The amount of loan outstanding captures all the loans from formal and informal sources, where informal sources include local moneylenders, friends and relatives. There are 1.4 million observations left after the cleaning process, which is described in detail in Appendix B.1.

Panel A in Table 1 shows that firms owned by members of historically disadvantaged castes (LC firms) are much smaller than others when size is measured by various variables. In terms of employment, LC firms account for 11.7% of total employment in the data, while HC firms account for 46.7%. In terms of gross output value, LC firms account for 8% of

[20] According to the Factories Act 1948: "Registration of manufacturing units is mandatory under Sectors 2m (i) and 2m (ii) of the Factories Act. Section 2m (i) refers to units engaging 10 or more workers and using power whereas 2m (ii) refers to units engaging 20 or more workers and not using power. Besides, some of the State Governments notify certain industrial activities for mandatory registration, although they do not conform to the criteria laid down under Sectors 2m (i) and 2m (ii). Such registrations are done under Section 85 (i) or Section 85 (ii) by the concerned State Governments. Section 85 (i) refers to units engaging less than 10 workers and using power and Section 85 (ii) refers to units engaging less than 20 workers and not using power."

the total gross output value in the data, while HC firms account for 61.3%. In terms of the material input value, LC firms account for 5.4% of the total material input value in the data, while HC firms account for 70.3%. In terms of total loans taken, LC firms account for 4.8% of total loans taken in the data, while HC firms account for 74.1%. Figure 1 plots the spatial variation in this asymmetry in entrepreneurial variables across caste, by plotting the absolute difference in log of gross value output between HC and LC firms. The plot shows a lot variation in this asymmetry, across districts in India.

Prices and quantities: The MSME census provides information on prices and quantities for three main *products* and three main *input materials* for the cross-section of firms during 2006-07 (see Appendix B.1 for more details).

Caste composition of employees: The MSME census also provides information on *employment composition by caste* for the cross-section of firms during 2006-07 (see Appendix B.1 for more details).

Retrospective panel 2004-2006-07: The MSME census does provide retrospective information on *gross sales* for the firms that survive up to 2006-07. This allows us to construct a balanced panel of MSMEs for the three-year period, 2004-05 to 2006-07. This synthetic panel allows us to compute statistics such as the *auto correlation of firms' output and input materials* (see Appendix B.1 for more details).

The main advantage of using the MSME data set is that it provides the caste of the enterprise owner, caste composition of employees, location of the firm, along with the firm balance sheet variables. Note that the MSME data set omits large enterprises (above a certain threshold of capital stock) unlike the ASI-NSS or the CMIE Prowess data sets. However, the latter data sets are not suitable for our analysis as they lack information on employee-employer castes, which is crucial for our study.

4.2 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.[21] The descriptive statistics are provided in Table 1. HC workers account for 28.3% of the employment survey sample, while LC workers account for 36.1% of the sample. However, Panel B in Table 1 shows that the HC workers report, on average, a higher daily wage (148.42 Indian rupees) compared to LC workers (59.47 Indian rupees), and a higher education.

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

4.3 Consumption Data

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

Panel B on individual wages and Panel C on household consumption in Table 1 together suggest that LC consumption is constrained and is much lower than HC consumption. This suggests that the *effective demand* faced by firms that cater to a larger fraction of the LC population are more likely to be bound by working-capital-constraints, and may not achieve their potential size. To investigate the causal nature of this correlation, we turn to the variation in monsoon rainfall across districts in India. The rainfall data is described next.

4.4 Rainfall Data

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

Cross-caste patterns from Tables Table 1 are consistent with the intuition that LC firms which cater to poorer households have stricter working-capital constraints and operate at a smaller size. To establish whether these correlations are causal in nature, consider the two sources of cross-sectional variation plotted in Figures 1 and 3. On the one hand, note that the cross-caste differences in economic conditions, across districts, are not coinciding with the spatial variation in rainfall. On the other hand, note that we are also able to track the unanticipated temporal deviations in rainfall, within districts. We will exploit this plausibly exogenous combination of socio-economic conditions of caste-groups across districts and temporal deviations in rainfall within districts, to provide causal evidence of the effect of caste-linkages on entrepreneurship in rural India.

5 Empirical Strategy

Motivated by the cross-caste patterns documented in Section 4, our empirical specification seeks to estimate the effects of caste-linkages on local firms' outcomes in India. We define a local economy as the economy of an Indian district.[22]

Our hypothesis for the following empirical exercises builds on the premise that historically disadvantaged community, denoted by LC, are poor and constrained (see Table 1). We posit that if caste-linkages determine firms' demand, the constraint on LC consumers should translate into stunted growth for LC firms. We shed light on this linkage by using an exogenous shift in local demand for LC households, due to higher rainfall, and observe its effects on firms owned by different caste-members across different sectors.[23] We assume that the effects of higher rainfall are transitory, that is, higher rainfall in a given year affects the local economy only in that year.[24]

5.1 Empirical Analysis

5.1.1 Effect of rainfall on individual wages, across caste

While previous work has already shown the average effect of rainfall on local wages, we motivate our setting by first providing evidence on the asymmetry in economic conditions across caste and the asymmetric effects of rainfall on income across caste.[25] This exercise helps us establish the baseline cross-caste differences in income levels and study the effect of a demand-shifter that especially shifts the income of LC individuals. The rationale for this income shift is as follows: higher rainfall within a district improves crop productivity and increases output and wages in the agricultural sector. Since LC workers are a crucial part of the agricultural labour force in rural India and have low income-levels, higher rainfall significantly improves their income.

In order to analyse the effect of higher rainfall on wages earned by workers across different castes, we estimate the following equation:

$$log(w_{it}) \;=\; \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_{dt} \times caste_i + \gamma X_{it} + \delta_d + \delta_t + \epsilon_{it} \tag{20}$$

where i denotes individual, d denotes the district, t denotes year. The regression includes three sets of fixed effects. First, we control for any time-invariant characteristics related to

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

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

[24] This is a standard assumption in prior work. See, for instance, Jayachandran (2006), Kaur (2019), and Santangelo (2019). In Section B.4, we also document the lack of serial correlation in monsoon rainfall, to rule out persistence in the occurrence of higher rainfall.

[25] See, for instance, Jayachandran (2006), Kaur (2019), and Santangelo (2019) for previous work on the effect of rainfall on wages of workers involved in the Indian agricultural sector.

caste that affect wages (e.g., occupational segregation within sector). Second, we control for any time-invariant district characteristics (e.g., main industries of employment or crops produced in a district). The third is time fixed effect to control for common trends across the country (e.g., sectoral national policies). We also include controls at the individual level: age, gender, education, land possessed, and crop season.[26] We also restrict our attention to workers between the age group of 18 and 60. In this and all regressions below, standard errors are clustered at the district level to capture serial correlation.[27] The coefficient $\beta_{3,LC}$ gives the elasticity of LC workers' wages to rainfall. According to our hypothesis, LC workers have lower income levels at baseline ($\beta_{2,LC} < 0$), but higher rainfall increases their income significantly ($\beta_{3,LC} > 0$), relative to HC workers.

5.1.2 Effect of rainfall on household consumption, across caste

The second step, in establishing the link between constrained LC households and LC firms, is to check whether the above patterns in income translate into a shift in demand. We hypothesise that LC households have relatively lower consumption levels, but the asymmetric benefits of higher rainfall especially relax the constraints of LC households and increase their consumption, through the effect of rainfall on agricultural wages. This helps us establish that higher rainfall can be interpreted as a demand-shifter for LC households. Using data on consumption patterns across product categories such as Manufacturing, Services, and Durables, we would also be able to investigate the effect of higher rainfall as a demand-shifter across non-agricultural product categories.

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

$$log(c_{ht}) = \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_{dt} \times caste_i + \delta_d + \delta_t + \epsilon_{it} \quad (21)$$

where h denotes household. The regression includes caste fixed effects and the two-way fixed effects (district and time) as in Equation (20). The coefficient $\beta_{3,LC}$ gives the elasticity of LC households' consumption to rainfall. According to our hypothesis, LC households have lower consumption levels at baseline ($\beta_{2,LC} < 0$), but higher rainfall increases their consumption significantly ($\beta_{3,LC} > 0$), relative HC households.

5.1.3 Effect of rainfall on firms, across caste

The third step, in establishing the link between constrained LC households and LC firms, is to check whether the above shift in LC households' demand particularly relaxes LC firms' cash-flow constraint and allows them to grow closer to their optimal size. This exercise helps

[26] We define a dummy equal to one if the month in which the daily wages of the worker is recorded is between June and December, and zero otherwise.

[27] See Section B.4 for an alternative specification of standard errors.

us establish the importance of caste-linkages in determining firms' demand in the local rural economy.

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

$$log(y_{ft}) \; = \; \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_d \times caste_i + \delta_d \times \delta_t + \delta_F + \epsilon_i \tag{22}$$

where f denotes the firm and F denotes the industry. We estimate the effect of higher rainfall on the following firm-level outcome variables (y): (1) revenue, (2) material input, and (3) ratio of revenue to material input. The regression includes district $\times$ year fixed effect to control for any time-varying district-specific characteristics. Note that this fixed effect subsumes the average effect of rainfall as well, along with district-specific trends (e.g., migration). We rely on the residual variation, that is the asymmetry in economic conditions across caste and further its asymmetric interaction with the deviations in rainfall across districts, to obtain a plausibly causal interpretation. We use the NIC product code (2004) at the 4-digit level to classify industries. The regression also includes a industry fixed effect to control for any product-specific characteristics common across the districts. The coefficient $\beta_{3,LC}$ provides the elasticity of LC firms' equilibrium outcomes to rainfall.

If a certain group of firms' demand is largely composed of constrained households' consumption, then these firms are more likely to exhibit stunted growth. We hypothesise that LC firms' demand is linked to LC households' consumption, and therefore, LC firms are more likely to be smaller. According to our hypothesis, we expect $\beta_{2,LC} < 0$ for outcomes (1) and (2). However, the lower size of LC firms could be a result of various factors. To identify the existence of caste-linkages in firm demand, we expect $\beta_{3,LC} > 0$ for outcomes (1) and (2). This is because an exogenous shift in LC households' demand translates into higher revenue for LC firms (who have a larger potential to grow) by increasing their cash-flow and relaxing their working-capital constraints, thus allowing them to grow closer to their optimal size.

5.2 Empirical Results

Individual wage elasticity: Table 1, Columns (1) to (3) present the results from estimating Equation (20), and show the asymmetric effect of higher rainfall on local wages. But first, note that the table shows the asymmetry in economic conditions across caste. LC agricultural workers earn 3% lower wages than HC agricultural workers, in a district with median rainfall. Second, wages of LC agricultural workers increase by 7.4% for every 1% increase in rainfall, which is 3.3% more than the increase for HC workers. This shows that workers across caste have asymmetric income levels and that higher rainfall induces a substantial transitory income change, particularly for the LC consumers in a district.

Household consumption elasticity: Table 1, Columns (4) to (7) present the results from

Table 1: Elasticities in rural India

	Individual wages			Household consumption				Firm outcomes	
	(1) All	(2) Agri.	(3) Non-Agri.	(4) All	(5) Mfg.	(6) Services	(7) Durables	(8) Revenue	(9) Material Input
log(rainfall)	0.050***	0.041*	0.069***	-0.005	-0.044	-0.002	-0.095	-	-
	(0.018)	(0.021)	(0.022)	(0.021)	(0.028)	(0.048)	(0.061)	-	-
MC	-0.028	-0.214*	0.158	-0.583***	-0.440***	-1.357***	-0.951**	-1.443***	-1.750***
	(0.130)	(0.123)	(0.136)	(0.147)	(0.138)	(0.380)	(0.415)	(0.308)	(0.417)
LC	-0.150	-0.250**	-0.043	-1.015***	-0.677***	-1.706***	-2.326***	-1.763***	-2.273***
	(0.114)	(0.105)	(0.134)	(0.138)	(0.162)	(0.343)	(0.377)	(0.314)	(0.400)
log(rainfall) $\times$ MC	-0.014	0.029	-0.040**	0.058***	0.042**	0.155***	0.113*	0.158***	0.191***
	(0.019)	(0.019)	(0.020)	(0.022)	(0.020)	(0.056)	(0.062)	(0.045)	(0.061)
log(rainfall) $\times$ LC	-0.001	0.033**	-0.015	0.100***	0.048**	0.138***	0.280***	0.172***	0.225***
	(0.017)	(0.016)	(0.020)	(0.021)	(0.024)	(0.051)	(0.056)	(0.046)	(0.059)
Observations	131,991	63,531	68,459	154,241	153,586	132,873	126,125	1,468,689	1,457,160
R-squared	0.401	0.353	0.394	0.351	0.251	0.225	0.218	0.422	0.460
Controls$_{it}$	✓	✓	✓	-	-	-	-	-	-
District FE	✓	✓	✓	✓	✓	✓	✓	-	-
Year FE	✓	✓	✓	✓	✓	✓	✓	-	-
District $\times$ Year FE	-	-	-	-	-	-	-	✓	✓
Product FE	-	-	-	-	-	-	-	✓	✓

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

estimating Equation (21), and show the asymmetric effect of higher rainfall on Monthly Per Capita Expenditure (MPCE). But first, note that the table shows the asymmetry in consumption behaviour across caste. LC households spend 35.5% lower than HC households per month, in a district with median rainfall. This difference is large in the services (79.5% lower) and durables (47.8% lower) sector. This shows that LC consumers are highly constrained. Second, LC households' consumption increase by 10%, for every 1% increase in rainfall, while HC households' consumption per month does not change significantly. A substantial fraction of this increase is explained by increase in spending in services and durables. This shows that households across caste have asymmetric consumption levels and that the rainfall-induced positive income shift to LC consumers translates into a substantial increase higher LC households' consumption.

Firm revenue elasticity: The above evidence shows that the higher rainfall induces an immediate effect on the local economy with a shift in demand for goods and services, largely driven by LC households. We now evaluate the resulting effect this has on firm outcomes, across different castes. Table 1, Columns (8) to (10) present the results from estimating Equation (22), and show the asymmetric effect of higher rainfall on firms' equilibrium outcomes.

But first, note that the table shows that LC firms are 62.8% smaller (by revenue) than HC firms, in a district with median rainfall. This is consistent with our hypothesis that LC firms are more likely to exhibit stunted growth, since they are subject to stronger constraints.

Second, corresponding to the increase in LC expenditure in Table 1, Columns (4) to (7), we see an increase in the revenue of LC firms by 17.2%, relative to HC firms, for every 1% increase in rainfall. Consistent with our hypothesis, this observation suggests that LC firms' constraints were relaxed due to a positive shift in demand from LC households, thus highlighting caste-linkages as a determinant of firm demand. This rise in LC firms' revenue is driven by the fact that constrained LC firms show an 22.5% rise in material input purchase, relative to HC firms, for every 1% increase in rainfall.

6 Quantitative analysis

In this section, We plan to estimate the parameters of the model using the data and the exogenous income shocks as described above. Finally, we will use this estimated model to perform various counterfactual experiments and ask questions such as how firm size distribution and income inequality would change if caste identity does not matter for product demand.

7 Conclusion

When standard market institutions are weak, as is the case in emerging economies, ethnic networks substitute in to assign roles to market participants, and provide structure to the economy. However, this also allows for the distortions present in the social fabric to permeate into the economic relationships. In this paper, we show caste-linkages in India affect product demand, such that firms face a segmented demand based on the caste of the firm owners. Firms are forced to reflect the constraints of their consumers, and experience stunted growth. Once those household constraints are relaxed, the firms are also able to grow closer to their optimal size.

Specifically, we show that historically disadvantaged (denoted by LC) communities earn lower wages and that their consumption is constrained, relative to others. Higher rainfall improves crop productivity and the wages of the LC communities involved in agriculture. Subsequently, we observe an increase in their consumption expenditure across a variety of products. However, this increase in demand is largely reflected as higher revenue only for firms belonging to the same communities.

This distortion in demand also has negative implications for the aggregate economy. It prevents the more productive LC firms from growing and reaching their optimal size. We document this using the dispersion in marginal product of material inputs ($mrpm$). We show that 1% higher rainfall can decrease the dispersion in *mrpm* by 5.3%, which is nearly one-third of the initial dispersion.

While our results suggest the presence of homophilic preferences as the cause of demand-side distortion, we are agnostic on the intensity of homophily or on the micro-foundation for such preferences (e.g., kinship, loyalty, information friction, geographical segregation). Such a study requires a richer data set and deserves further research.

References

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

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

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

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

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.

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.

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

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

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

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

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

DESCHÊNES, O. AND M. GREENSTONE (2007): "The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather," *American economic review*, 97, 354–385.

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

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

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

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2016): "The slow growth of new plants: Learning about demand?" *Economica*, 83, 91–129.

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

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

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

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

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

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

HSIEH, C.-T., 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 (2014): “The life cycle of plants in India and Mexico,” *The Quarterly Journal of Economics*, 129, 1035–1084.

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

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

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

MSME (2009): *Fourth All India Census of Micro, Small and Medium Enterprises, 2006*, Ministry of Micro, Small and Medium Enterprises.

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

MUNSHI, K. AND M. ROSENZWEIG (2016): “Networks and misallocation: Insurance, migration, and the rural-urban wage gap,” *American Economic Review*, 106, 46–98.

OH, S. (2021): “Does identity affect labor supply?” *Available at SSRN 3998025*.

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

TYBOUT, J. R. (2000): “Manufacturing firms in developing countries: How well do they do, and why?” *Journal of Economic literature*, 38, 11–44.

A Model Derivations

A.1 Household Problem

The first-order conditions for the households are

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

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

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

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

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

A.2 Firm Problem

A.2.1 Proof Proposition 1: Optimal Location

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

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

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

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

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

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

Collecting terms and a few steps of Algebra give us

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

where $B_{s'} = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\mathcal{C}_{s'}\right)^{1-\sigma} z^{\sigma}$ and $\mathrm{Y} = \sum_{s\in\mathcal{S}}\kappa_s$, and $\lambda_{ss',k} = \frac{\beta\kappa_s + B^{-1}_{s'}\gamma_{s',k}\mathbb{1}_{s=s'}}{\mathrm{Y}}$. $\mathbb{1}_{s=s'}$ is the indicator function that have value one if $s=s'$. Under these definitions, we can rewrite the profit maximisation as reduced down to an optimal (minisum Fermat-Weber Problem) location problem.[28]

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

First-order conditions are given by

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

This gives us the expression for the optimal distance moved

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

A.2.2 Taste Shifters and Transportation Costs.

The revenues are given by

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

Total firm-level profits can be written as

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

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

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

We assume $\Psi(d_{ss'},\Delta d_{s'}) = e^{-\hat{\beta}\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s',k})^2\right)}$ and $\tau(d_{ss'},\Delta d_{s'}) = e^{\gamma\left(\sum_{k\in\mathcal{N}}(d_{ss',k}-\Delta\mathcal{X}_{s',k})^2\right)}$ and de-

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

fine $\tilde{\beta} = \hat{\beta}\sigma + \gamma(\sigma - 1)$. We assume the first-order Taylor approximation of

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

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

A.2.3 A side note on Pareto

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

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

A.2.4 A side note on Price index

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

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

A.2.5 Derivation of Firm-level Revenues

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

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

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

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

A.2.6 Proof of Proposition 2

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

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

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

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

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

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

Taking the derivative both of the revenue equation, we get

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

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

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

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

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

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

A.2.7 Proof of Proposition 3

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

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

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

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

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

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

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

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

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

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

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

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

B Empirical Analysis

B.1 Data

B.1.1 MSME

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

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

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

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

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

Winsorization: The financial variable such as market value of fixed assets, gross value-added, total wage-bill, employment, amount of loan-outstanding, gross output, total cost of variable input and net-worth are winsorized at 1 and 99th percentile. Furthermore, the variables used in regressions, $mrpm$ and $mrpl$, are winsorized at 1 and 99th percentile.

Table 1: Summary statistics

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

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

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

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

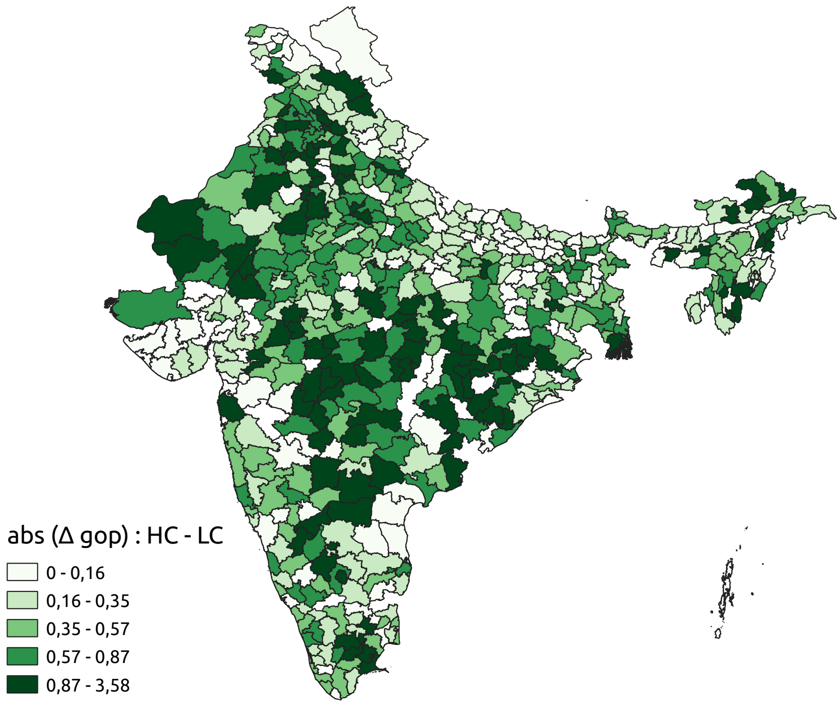


Figure 1: Spatial variation in output-difference between HC and LC firms, at the district-level in India (2006-07)

Notes: This figure uses MSME 2006-07 data to plot the absolute difference in log(gross output value) between firms owned by members historically classified as high-caste and firms owned by members historically classified as low-caste.

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

Panel B in Table 2 shows the caste-wise employment for firms belonging to different castes. We see a very distinct positive assortative employer-employee matching across castes. This further enhances the notion of ethnic identity of firms, which we parameterise in our quantitative model.

The MSME datset also provides retrospective information on sales and input materials for the enterprises that survive up to 2006-07. This allows us to construct a balanced panel of

Table 2: MSME: Additional summary statistics

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

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

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

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

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

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

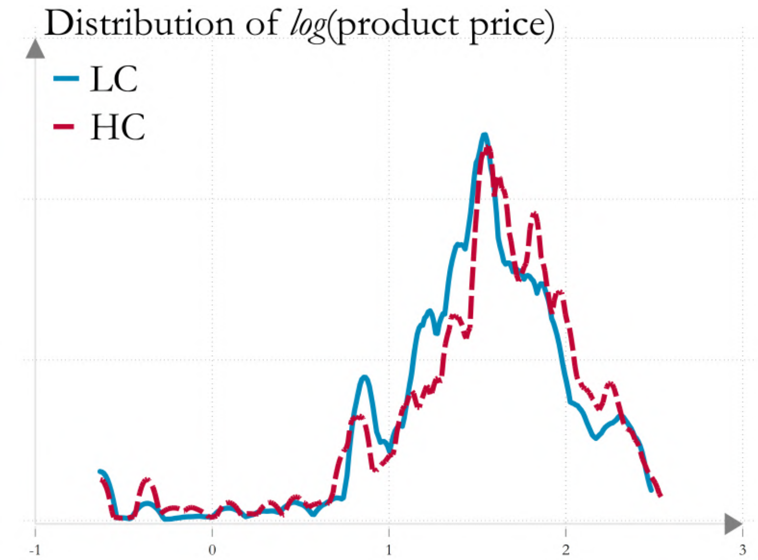
Notes: The table reports descriptive statistics. LC, MC, and HC represent entrepreneurs historically classified as belonging to the historically disadvantaged castes, middle-castes, and historically privileged castes, respectively. Sampling multipliers are applied in all the panels.
Panel A reports statistics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

MSMEs for the three-year period, 2004 - 2007. This panel information allows us to compute statistics such as auto-correlation in output and materials. Panel C in Table 2 shows that the auto-correlation in output is much higher for LC firms as they struggle to grow, while the auto-correlation in input materials is similar across castes.

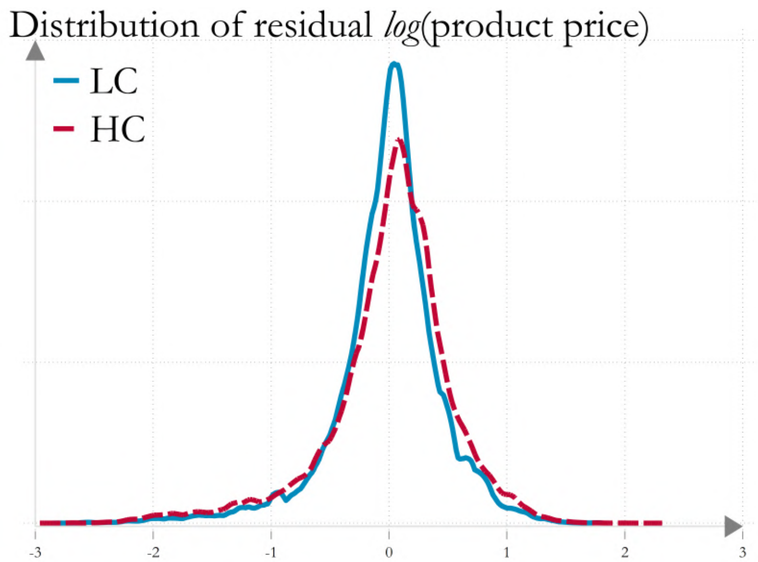
B.1.2 Rainfall

The Tropical Rainfall Measuring Mission (TRMM) provides gridded rainfall rates at very high spatial and temporal resolution. Daily rainfall measures are available at the 0.25 by 0.25 degree grid-cell size. Spatially, we aggregate this data by calculating the mean rainfall registered on the grid points within the boundary of a district. Temporally, we aggregate this data as the total rainfall during the months of June, July, August and September. Together, we obtain a measure of rainfall at the district-year level.[29] For Indian districts between 1980

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



(a) unconditional



(b) within District and Product-Category

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

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

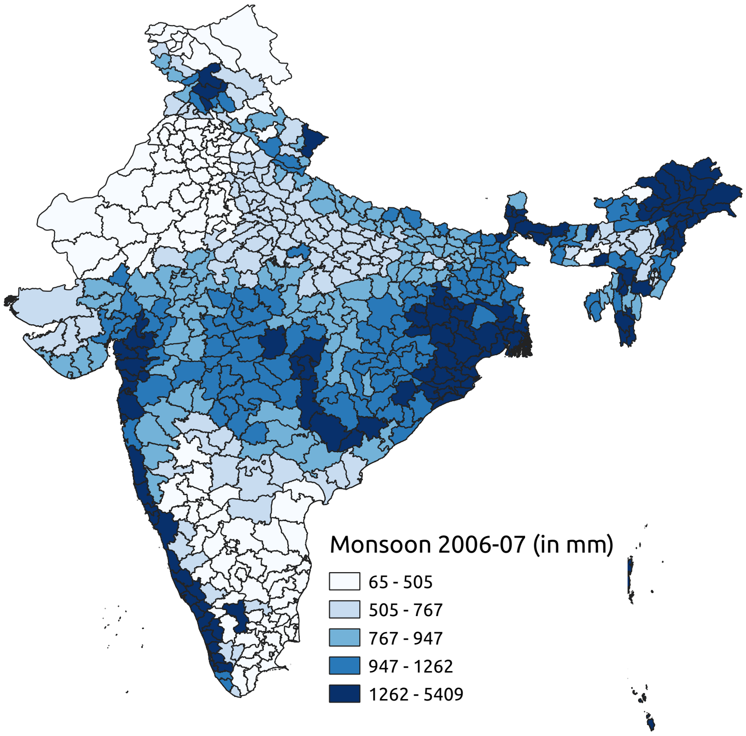


Figure 3: Spatial variation in monsoon rainfall, at the district-level in India (2006-07)
Notes: This figure uses TRMM 2006-07 data to plot the total monsoon rainfall received by Indian districts, in millimeters of rainfall.

and 2016, monsoon rainfall accounts for more than 75% of annual rainfall[30]. Figure 3 plots the spatial variation in rainfall, measured in millimeters of monsoon rainfall.

B.2 Geographical segregation

The premise of our study is to investigate whether firms belonging to a certain ethnic group tend to cater to consumers of the same group, and hence suffer from the consequences to growth due to limited demand. In Section 5, we have documented the presence of such ethnic linkages and the resulting segregation in the product market. This overall segregation may be a combination of two underlying channels: (a) homophilic preferences, where consumers have a preference to buy from firms belonging to ethnically similar groups, and (b) geographical distance, where consumers have a cost to access products that geographically farther.[31]

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

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

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

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

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

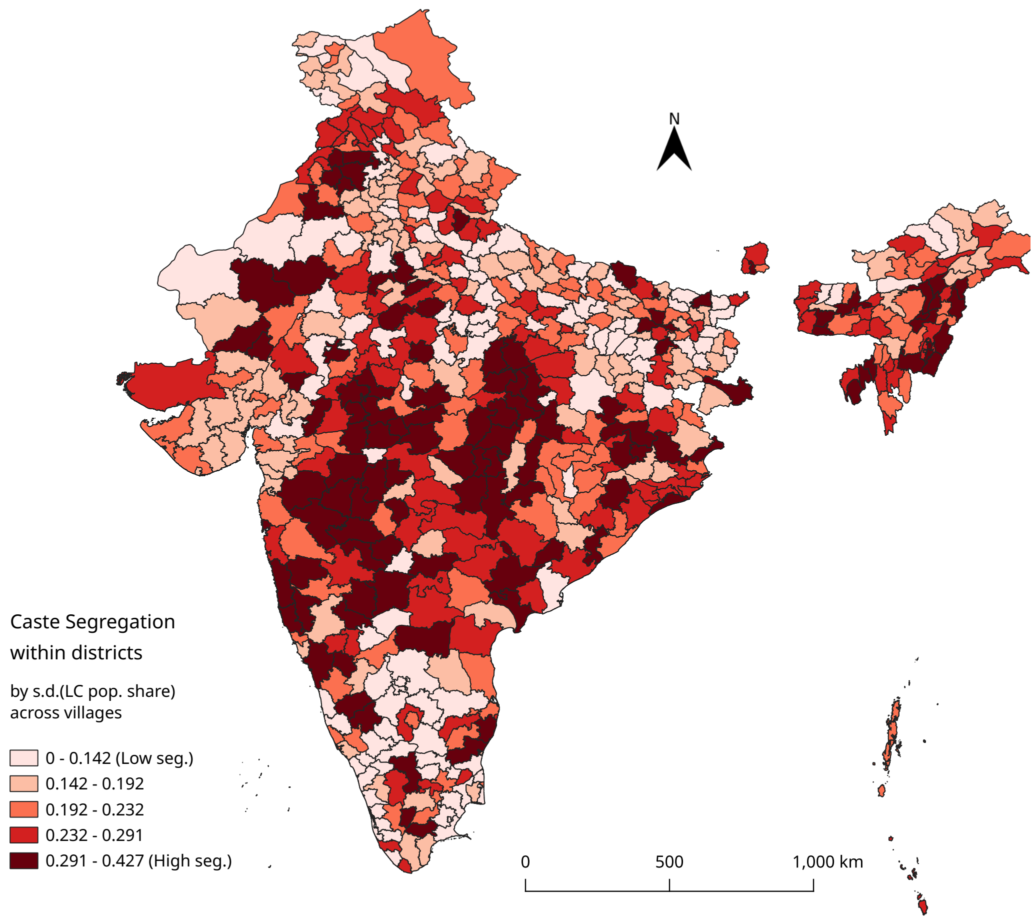
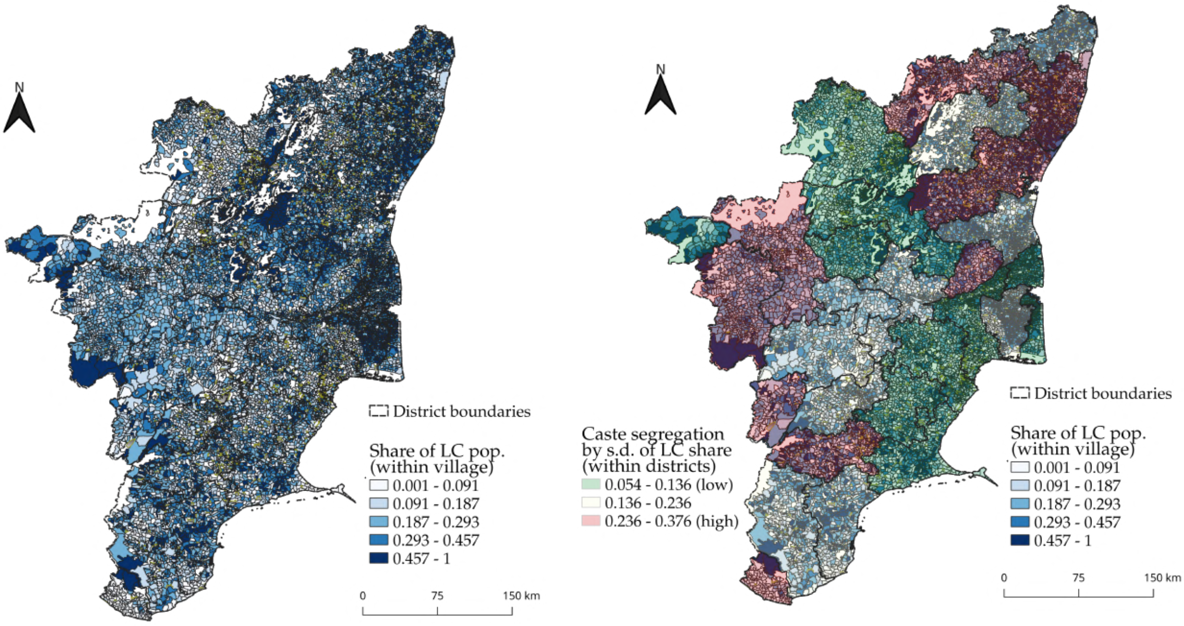


Figure 4: Geographical segregation, by caste, across districts of India (2001)
Notes: This figure uses Socioeconomic High-resolution Rural-Urban Geographic Data Platform for India (SHRUG) data's Economic and Population Census module (2001) to plot the standard deviation in the share of LC (SC+ST) population across villages within each district across India.



(a) share of LC population across TN villages

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

Figure 5: Geographical segregation, by caste, across districts of Tamil Nadu (2001)

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

Table 3: Elasticities in *unsegregated* rural India

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

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

Figure 4 shows that there is a lot of variation in within-district geographical segregation across India. Figure 5 provides a more granular look at this measure by focusing on the state of Tamil Nadu. Having constructed this measure, we use its spatial variation and replicate our firm-level empirical analysis. We omit observations of highly *geographically segregated* districts, that is those districts that belong to the top quartile of our measure of segregation. Table 3 shows that the results are in line with our baseline results from Table 1.

This evidence suggests that geographical segregation may not be the driving factor behind the homophilic demand patterns observed in Table 1. However, note that the interaction coefficients for LC firms in Columns (8) and (9) are larger than the baseline, suggesting that in districts with low geographical frictions, stronger homophilic patters can lead to larger revenues for LC firms. Therefore, we incorporate the presence of geographical frictions to obtain a more complete picture of the Indian rural economy and parameterise the forces due to it in our quantitative model (see Table **??** for more details).

B.3 Additional evidence

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

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

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

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

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

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

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

Table 4: Cross-sectional firm-level elasticities in rural India: MSME 2006-07

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

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

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

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

Labour. In Table 4, Columns (1) and (2), we begin by establishing that the results from the MSME panel shown in Table 1 hold in the cross-section as well. Secondly, Table 4 shows that wages paid by LC firms also increase by 7.1% relative to HC firms, following higher rainfall, consistent with their constraints being relaxed and the material input purchase increasing. However, the table shows that the constraint of LC firms was especially relaxed with regards to their material input purchase than wages. This suggests that with increased demand, the constrained LC firms were able to obtain more *variable input* and necessary workers in order to expand.

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

Credit. Thirdly, we investigate whether the positive income shift for LC households was transferred to LC firms through channels other than a shift in demand, namely that of loans. First, table 4 shows that value of loans LC firms have taken is 68.5% lower than that of HC firms, in a district with median rainfall.[32] Second, with higher rainfall, there is no change in loans taken overall, from formal or informal institutions. This rules out any credit-side channel through which the LC income shift reaches firms, that is through formal (e.g., banks) or informal (e.g., caste networks) credit sources. An alternative outcome that may capture the generation of credit is the marginal revenue product of capital ($mrpk$). Any fall in this measure would suggest that firms have obtained funding (external or internal) and have invested in capital after rainfall. Table 4, Column (9) shows no significant change along these lines. This evidence provides further support in the identification of cash-flow based constraints among firms in rural India.

Labour composition. Our model builds a framework of firm identity which is determined by both employee composition and firm owner's identity, to analyse the effects of caste-based linkages in consumer demand. To further understand the increase in size of LC firms and the caste-linkages in employment, we decompose the change in the labour composition of firms by caste. Table 5 shows that LC firms' increase in employment is driven by an increase in own-caste hiring. This suggests that LC firms increased size while sustaining their LC identity to cater to the LC population-driven local positive demand shock.

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

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

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

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

Table 5: Labour composition elasticity in rural India: NSS 2004-10

	(1) Total Wages	(2) Total Workers	(3) Total HC Workers	(4) Share of HC Workers
MC	-0.873***	-0.638***	-0.429***	0.233**
	(0.231)	(0.188)	(0.140)	(0.113)
LC	-0.890***	-1.008***	-0.692***	0.336**
	(0.254)	(0.260)	(0.207)	(0.130)
log(rainfall) $\times$ MC	0.096***	0.109***	0.042**	-0.070***
	(0.033)	(0.027)	(0.020)	(0.017)
log(rainfall) $\times$ LC	0.070*	0.145***	0.081***	-0.067***
	(0.036)	(0.037)	(0.030)	(0.019)
Observations	483,151	263,051	263,051	263,051
R-squared	0.471	0.385	0.301	0.260
District FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes. The regressions of (logarithmic) (i) total wages, (ii) total workers, (iii) total HC workers, and (iv) share of HC workers, on rainfall (logarithmic) for the full sample of workers. The additional control variables are age, gender, education, land possessed, and crop season. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

B.4 Robustness Checks

B.4.1 Robustness of wage patterns to worker sample

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

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

Second, we check for robustness of the effect of rainfall, across caste, to other determinants of household consumption patterns. Using land ownership as a proxy for within caste variation in households' illiquid wealth status, we analyse their consumption patterns by implementing a triple interaction exercise. This exercise addresses concerns regarding the reliance of our results to the segmentation of demand along the lines of wealth, instead of caste. The results reported in Table 8 are consistent with the baseline results from Table 1, Columns (4) to (7). It further shows that LC consumers with less land, who would be more constrained, spend higher than LC consumers with high land. This shows that caste is an important determinant of consumption patters in the local economy, and within-caste wealth inequality is an additional determinant.

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

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

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

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

In Table 9, we interact rainfall with educational qualification of the head of household, and with the number of meals per day consumed by the head of the household. In rural India, head of the household is responsible for a large portion of the household's liquid income. This exercise can help us understand whether the caste-driven consumption patterns documented in Table 1, Columns (4) to (7) can instead be explained by consumption patterns common to poor households. But this is not the case, as we see that the results are qualitatively similar to the baseline Table 1, Columns (4) to (7).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

In Table 11, we restrict our attention to industries where the mean difference between the prices of LC and HC firms' products are less than 0.1 standard deviations. This exercises addresses concerns on whether the demand for low-quality/priced products from LC households is the mechanism behind our resutls. We find that firms' outcomes are qualitatively similar to the baseline Table 4. This evidence adds support to our claim that caste-based consumer-firm linkages is key to understanding firm growth in rural India.

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

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

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

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

In our baseline empirical analysis, presented in Table 1, we measure the ethnic identity of the firm using the caste of the owner of the firm. It is also possible that the caste composition of workers (e.g., salespersons) employed in firms determine the perception of ethnic identity among consumers. In this exercise, we use the share of non-HC employees working in a firm as a measure of firms' ethnic identity. Table 12 shows that firms which had a larger population of non-HC employees exhibited larger growth compared to other firms, after higher rainfall. This evidence is in line with our findings in our baseline analysis where after a higher rainfall consumers of LC and MC communities received a positive income shock and increased their expenditure on firms ethnically closer to their identity.

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

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

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

B.4.7 Robustness to alternative clustering of standard errors

In Table 13, we cluster standard errors at the state-year level to allow for spatial correlation. The results are qualitatively consistent with our baseline results from Table 1.

B.4.8 Robustness to alternative definition of rainfall

In Table 14, we use an alternative definition of rainfall. We follow Jayachandran (2006) to define a categorical variable aimed at capturing possible non-linearities in the effects of rainfall. This dummy variable equals one if monsoon rainfall is greater than the district's eightieth percentile of monsoon rainfall, zero if between the twentieth and eightieth percentiles, and minus one if below the twentieth percentile. The percentiles are calculated using the last twenty years data. The results are qualitatively consistent with our baseline results from Table 1.

B.4.9 Evidence on the transitory nature of rainfall deviations

Table 15 shows that there is no serial correlation in rainfall across districts. This adds support to the interpretation of the exogenous variation in rainfall as having an unanticipated and transitory effect on the local economy.

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

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

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

Table 14: Elasticities in rural India, alternative definition of rainfall

	Individual wages			Household consumption				Firm outcomes	
	(1) All	(2) Agri.	(3) Mfg.	(4) All	(5) Mfg.	(6) Services	(7) Durables	(8) Revenue	(9) Material Input
rainfall	-0.004	0.002	0.015	-0.006	-0.017	-0.020	-0.100**	-	-
	(0.015)	(0.015)	(0.020)	(0.017)	(0.020)	(0.038)	(0.050)	-	-
MC	-0.121***	-0.025*	-0.105***	-0.203***	-0.158***	-0.348***	-0.201***	-0.385***	-0.468***
	(0.012)	(0.013)	(0.014)	(0.016)	(0.016)	(0.041)	(0.038)	(0.032)	(0.041)
LC	-0.155***	-0.030**	-0.139***	-0.357***	-0.360***	-0.801***	-0.504***	-0.613***	-0.769***
	(0.013)	(0.013)	(0.014)	(0.016)	(0.019)	(0.039)	(0.040)	(0.033)	(0.042)
rainfall × MC	0.019	0.024	0.001	0.037	0.009	0.103*	0.023	0.007	0.032
	(0.017)	(0.017)	(0.024)	(0.023)	(0.023)	(0.057)	(0.054)	(0.037)	(0.048)
rainfall × LC	0.040**	0.034**	0.026	0.059***	0.032	0.088*	0.210***	0.009	0.062
	(0.017)	(0.016)	(0.023)	(0.022)	(0.026)	(0.052)	(0.059)	(0.047)	(0.061)
Observations	131,991	63,531	68,459	154,241	153,586	132,873	126,125	1,468,689	1,457,160
R-squared	0.401	0.352	0.394	0.349	0.251	0.224	0.216	0.421	0.459
Controls$_{it}$	✓	✓	✓	-	-	-	-	-	-
District FE	✓	✓	✓	✓	✓	✓	✓	-	-
Year FE	✓	✓	✓	✓	✓	✓	✓	-	-
District × Year FE	-	-	-	-	-	-	-	✓	✓
Product FE	-	-	-	-	-	-	-	✓	✓

Notes. In Columns (1) to (3), the regressions are of individual-level daily wage (logarithmic) on rainfall, a categorical variable equal to one if monsoon rainfall is greater than the district's eightieth percentile of monsoon rainfall, zero if between the twentieth and eightieth percentiles, and minus one if below the twentieth percentile. The percentiles are calculated using the last twenty years data. *Agri.* stands for agricultural workers in rural areas. $Non - Agri.$ stands for workers employed in sectors non-agricultural sectors such as manufacturing, constructions, services, etc. in rural areas. The additional individual-level control variables included are age, gender, education, land possessed, and crop season. Sample includes individuals between the ages of 18 and 60. In Columns (4) to (7), the regressions are of households' monthly per capita consumption (logarithmic) on rainfall (logarithmic). $Mfg.$ stands for consumption of clothing and footwear. In Columns (8) to (10), the regressions are of firm-level variables (logarithmic) on rainfall (logarithmic). Sampling multipliers are applied in all regressions. Standard errors in parentheses are clustered at district level in all regressions, *** p<0.01, ** p<0.05, * p<0.1.

Table 15: Testing for serial correlation in rainfall

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

Notes. This table tests for serial correlation in rainfall.

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

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