Product Demand, Misallocation and Firm Growth*

Sampreet S. Goraya† Akhil Ilango‡

February 28, 2023

This paper examines the impact of ethnic networks on product demand and firm growth 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 communities (LC) experience a positive shift in income due to good rainfall, relative to others, they translate it into higher per-capita expenditure across various non-agricultural product categories, whose supply is not directly affected by rainfall. Second, we exploit this demand-shifter to document how product demand is segmented along caste lines. This demand-side distortion creates input and productivity dispersion, across firms owned by different caste-members. Specifically, the increase in LC demand, due to higher rainfall, leads to higher revenue for the firms owned by members of the same caste category (LC firms), relative to others. In terms of input misallocation, LC firms exhibit 16.1% higher $mrpm$ (marginal revenue product of variable input) relative to other firms, in a district with median rainfall. An increase in rainfall by 1% reduces this $mrpm$ dispersion between LC and HC firms by 5.3%, due to a rise in LC demand and, subsequently, LC firms' size. We rationalize these observations using a framework of heterogeneous working-capital constraints for firms. These results suggest that, in the short run, input misallocation and stunted growth for firms are caused by limited demand, rather than the other way around.

Keywords: Caste System, Misallocation, Firm growth, Demand-side constraints.
JEL Codes: O11, D43, D61.

*We thank Rodrigo Carril, Rosa Ferrer, Sandro Shelegia and audiences at CREI international lunch, CREI macro lunch for helpful discussions and comments.

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

‡Department of Economics, Universitat Pompeu Fabra and Barcelona GSE. Email: akhil.ilango@upf.edu

1 Introduction

Local demand is an important determinant of firm growth in emerging economies.[1] Since these economies suffer from weak formal institutions, firms rely on informal institutions, such as the social structure of caste in India, to solve their incentive and commitment problems.[2] These informal institutions can generate distortions, stunt the growth of firms, and hamper the aggregate productivity of the economy. In this paper, we propose and empirically document a novel channel of distortion, arising from such an informal institution. Specifically, using new data on the caste of firm owners, we study the effect of caste system in India on firm growth and aggregate productivity, through the channel of demand segmentation - that is, the demand for goods produced by a firm owned by a member of a certain caste is primarily driven by the households of the same caste.

First, we document that historically disadvantaged individuals belonging to the lower categories of the caste hierarchy (hereafter denoted by LC) earn lower wages and consume less, relative to historically privileged individuals belonging to the higher categories of caste hierarchy (hereafter denoted by HC). Further, firms owned by members of LC (hereafter referred to as LC firms) are smaller in size, but operate at a higher *marginal revenue product of material inputs* (our measure of productivity), relative to the firms owned members of HC (hereafter referred to as HC firms). This suggests a differential barrier to the growth of LC firms, relative to HC firms. In our second string of evidences, we show that when there is an exogenous positive shift in income to LC households, they respond by increasing their expenditure across various product categories. Subsequently, within product categories, the LC firms exhibit a substantial increase in revenue, relative to HC firms. These evidences suggest that caste-linkages are an important driver of firms' demand and hence, firm growth, in rural India. We further find that this increase in firm size is not driven by a change in prices or access to credit.

Using a framework of heterogeneous cash-flow constraints for firms, we provide closed-form analytical solutions that, first, rationalises the above findings on the effect of an asymmetric shift in demand on firm growth. Second, we use our theoretical framework to motivate our measure of productivity and derive testable implications for the effect of demand-side distortions on the aggregate productivity of the economy. Our model allows us to illustrate a new channel within a well-established macroeconomic framework.[3] The intuition behind our findings is as follows: Since LC households are constrained,

[1] See, for instance, Jensen and Miller (2018) for recent evidence on the limits of market integration in the South Indian fisheries sector.

[2] See, for instance, Boehm and Oberfield (2020) for recent evidence on the effect of weak enforcement on misallocation in India. See Munshi (2019) for an overview of the role of caste in the various facets of the Indian economy.

[3] See, for instance, Drechsel (2022) for some recent work on cash-flow based constraints and its macroeconomic implications.

LC firms are also working-capital constrained and purchase less material input ex-ante, relative to HC firms. Due to such caste-linkages in consumption, when there is a shift in LC households' income, it translates into higher demand for LC firms, relative to HC firms. Once there is a positive shift in the demand LC firms face, it allows them to purchase more material input, and leads to an increase in production and revenue, relative to other firms. Overall, this shift in demand results in a decline in the dispersion in marginal revenue product of material inputs across firms, and an increase in the aggregate productivity of the economy.

To analyse 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 asses 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 shift 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.[4] Higher rainfall increases agricultural productivity and therefore, wages 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.

Next, to analyse the effect of the shift in income on consumption patterns across caste, 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. Motivated by this, we next investigate the indirect effects that higher rainfall causes on firms' input, 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 monsoon 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. As constrained face more demand with higher rainfall, LC firms' are now able to purchase more material input, thus decreasing input dispersion,

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

and are move closer to their optimal size than earlier. As these productive LC firms are now able to grow, this also reduces the dispsersion in productivity in the local economy: we observe a 5.3% decrease in marginal product of material input, for a 1% increase in rainfall.

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 see no change. Considering input prices as a proxy for product quality, this result reassures us that changes in product quality is unable explain the stylized facts. Further, no change in prices also shows that the rise in revenue is driven by rise in quantity sold or growth in the size of the firms. 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 suggest that the demand-side constraints are a crucial driver for the growth of LC firms in rural India.

1.1 Literature Review

Our paper contributes to two broad strands of the literature on firm-level misallocation, and on the demand-side determinants of firm growth.

First, our paper contributes to the literature on race and ethnicity-based distortions, and misallocation. Munshi and Rosenzweig (2016) and Oh (2021) study the distortions generated by caste and Hsieh, Hurst, Jones, and Klenow (2019) and Cassan, Keniston, and Kleineberg (2021) study the distortions generated by race in allocation of talent across space and occupations. Hjort (2014) shows that distaste between ethnicities distorts the input market and generates misallocation in a Kenyan flower market. Banerjee and Munshi (2004) document the misallocation of capital across communities in Tirupur (India) and Goraya (2023) studies the aggregate asymmetry in access to capital for firms owned by members of different caste. While the above papers study distortions that affect the input market directly, our paper complements this literature by documenting a new source of distortion that affects the output market - asymmetric cash-flow constraints for firms due to caste-based demand - and its implications for allocation of resources among firms.

Second, our paper also adds to the growing body of evidence on the importance of demand-side factors for firm growth. Foster, Haltiwanger, and Syverson (2016) show 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) use bar-code data to show that "firm appeal" explains a lot of the firm size heterogeneity in the US. Einav, Klenow, Levin, and Murciano-Goroff (2021) use customer transactions 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) show that the number

of final buyers explain 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 intensive margin of demand, especially when the extensive margin faces resistance to change (due to caste in our setting). Our paper provides further evidence that even when firms are more productive than others, a limited and constrained customer base will lead the firms to suffer from stunted growth.

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 cash-flow based financial constraints that affect firms heterogeneously and derives testable equilibrium implications. Section 4 introduces the data, 5 discusses our empirical strategy, and Section 6 documents the empirical evidence. Section 7 concludes by discussing some avenues for future research. Proofs are provided in Appendix A.

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.[5] The hierarchy dictated by the system, based on historical occupation, in descending order of dominance, is as follows: Brahmins (priests and teachers), Kshatriyas (rulers and soldiers), Vaishyas (merchants and traders), and the Sudras (laborers and artisans). Further, there are two additional groups that fall outside the caste system. The first one embodies the group of people traditionally known as Dalits.[6] 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.[7] 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

[5] 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.

[6] 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.

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

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." HC workers account for 33.3% of the employment survey sample, while LC workers account for 28.9% of the sample. However, the HC workers report, on average, a higher daily wage (84 Indian rupees) compared to LC workers (48 Indian rupees), and a lower employment share in agriculture (28% of HC workers compared to 42% of LC workers are involved in the agricultural sector).

3 A Theoretical Framework

We model a local rural economy with heterogeneous consumers, product differentiation and endogenous firm size distribution, to guide the empirical analysis. Firms in this economy face heterogeneous cash-flow constraint. This is a stylized approach to capture the heterogeneity in firms' composition of demand, based on the caste of the firm owner. Firms that cater to a larger share of poorer households are constrained from reaching their optimal size, since they have lower working-capital. Due to barriers in access to institutional credit, working capital continues to be an important source of credit for firms in rural India.[8]

We use this framework to develop testable implications of the effect of demand-side distortions on firms' growth in developing countries, where informal institutions such as caste play a crucial role in determining market outcomes. Specifically, we will draw implications for how a positive shift in income of LC households affects outcomes of firms.

The key channel that we study is as follows: heterogeneous consumers make decisions on how much and where to consume from, which asymmetrically affects firms' cash-flow constraint, and hence, the equilibrium firm size distribution and aggregate misallocation.[9]

[8] See, for instance, Goraya (2023) for evidence on barriers in access to credit to firms owned by individuals belonging to historically disadvantaged groups in India.

[9] See, for instance, Lian and Ma (2021) for recent empirical evidence that 80% of corporate debt is cash-flow based, compared to asset-based, for firms from the non-financial sector in the US. Also, see Drechsel (2022) for macro-implications of cash-flow constraints for US firms.

3.1 Consumers' Preferences

A representative consumer's utility, U, is captured using a standard CES utility function

$$\begin{aligned} U &= \left[\sum_j q_j c_j^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \\ s.t.\ I &= \sum_j p_j c_j \end{aligned}$$

where σ denotes the elasticity of substitution between goods $j \in \{1, ...N\}$, c_j denotes the units of consumption of good j, q_j denotes the utility derived per unit of good j consumed, p_j denotes the price of good j, and I denotes the income.

Deriving the demand for good j, we get an iso-elastic residual demand curve

$$c_j = q_j^{\sigma} p_j^{-\sigma} \frac{I}{\mathcal{P}^{1-\sigma}}$$

where $\mathcal{P} = \left[\sum_j q_j^{\sigma} p_j^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$ denotes the aggregate price index. Let us define $d_j = q_j^{\sigma} \frac{I}{\mathcal{P}^{1-\sigma}}$ as *effective demand*. Any positive shift to the income, I, will translate into higher demand for all goods in this representative consumer's consumption bundle. In other words, it causes a shift in d_j. If this bundle primarily consists of the goods produced by LC firms, then the income shift will essentially act as a demand shifter for these firms. If y_j is the total output of firm j and if there is no excess supply of goods then $c_j = y_j = p_j^{-\sigma} d_j$, and $p_j = d_j^{1/\sigma} y_j^{-1/\sigma}$.

3.2 Production Sector

There is a continuum of firms $j \in (0, 1]$ with a total mass $\mathcal{F}$. Each firm produces a differentiated variety. Firms have access to a Cobb-Douglas production technology with constant returns to scale. Firms use two production inputs: own productivity z_j and total variable input v_j. In this case, z_j also reflects the difference in investment in plant and machinery. The production function is given by

$$y_j = z_j v_j.$$

The revenue of the firm is denoted by $r_j = p_j y_j$. Using the downward sloping curve from the household problem, one can rewrite revenue as $r_j = d_j^{\frac{1}{\sigma}} y_j^{\frac{\sigma-1}{\sigma}}$.

3.2.1 Unconstrained firms

As a benchmark, consider the familiar unconstrained problem. We normalize the price of *variable input* to one.

$$\max_{v} \pi(z, v) = r(z, v) - v \tag{1}$$

Its solution is given by

$$v^* = \left(\frac{1}{\mu}\right)^{\sigma} d_j z_j^{\sigma-1}; \qquad r^* = \left(\frac{1}{\mu}\right)^{\frac{\sigma}{\mu}} d_j z_j^{\frac{\sigma}{\mu}}; \qquad mrpm_j^* = \frac{\sigma - 1}{\sigma} = \frac{1}{\mu} \tag{2}$$

where $\mu = \frac{\sigma}{\sigma-1}$ denotes firm markup, and $mrpm_j^*$ denotes the unconstrained *marginal product of material input*. Equation (2) shows that the *effective demand*, d_j, determines firm size (increases linearly) but does not affect the $mrpm_j^*$ for unconstrained firms.

Since the price of variable input, p^v, is normalized to one, $mrpm$(= *marginal product of m*× p^v) gives the wedge in the price of variable input. For unconstrained firms, we see that it is a constant and there is no dispersion. Therefore, any dispersion in $mrpm$ observed in the data is evidence for dispersion in effective price (net of constraints/frictions) of variable input across firms.

Further, note that any shift in *effective demand*, d_j, does not affect $mrpm$ across unconstrained firms. Therefore, weaker working-capital, depending on a firm's *effective demand*, may introduce heterogeneous wedges in the use of input, and hence allow us to interpret $mrpm$ as a measure of input misallocation.[10] We pursue this formally below.

3.2.2 Constrained firms

Now, let us consider the case of constrained firms, which need to finance their expenditure on input by their cash-flow. Hence, they rely on their *effective demand*. In other words, constrained firms face working-capital constraints such that

$$v_j \;\leq\; \psi(r_j) r_j, \tag{3}$$

where $\psi(r_j)$ is a parameter used to capture the degree of financial friction (e.g., collateralizability of its effective demand) a firm faces. We assume that $\frac{\partial \psi(r_j)}{\partial r_j} \geq 0$, such that large firms face a lower degree of financial friction, relative to small firms.

For simplicity, we consider the functional form $\psi(r_j) = \lambda r_j^{\phi}$, where $\lambda, \phi \geq 0$. Now, one can rewrite the working capital constraints as $v_j \;\leq\; \lambda r_j^{1+\phi}$. Using Equation (3), we obtain

[10] A commonly used approach to quantify the role of factor market frictions, pioneered by Hsieh and Klenow (2009), shows that, under some conditions, frictions translate into wedges in firms' marginal products of input. Therefore, evidence of dispersion in $mrpm$ can be inferred as the amount of misallocation.

the optimal size of the constrained firms as follows

$$v_j = \lambda_1 d_j^{\frac{(1+\phi)\mu}{\sigma(\mu-1-\phi)}} z_j^{\frac{1+\phi}{\mu-1-\phi}}; \quad r_j = \lambda_2 d_j^{\frac{\mu}{\sigma(\mu-1-\phi)}} z_j^{\frac{1}{\mu-1-\phi}}; \quad mrpm_j = \lambda_3 d_j^{\frac{-\mu\phi}{\sigma(\mu-1-\phi)}} z_j^{\frac{-\phi}{\mu-1-\phi}}, \tag{4}$$

where $\lambda_1 = \lambda^{\frac{\mu}{\mu-1-\phi}}$, $\lambda_2 = \lambda^{\frac{1}{\mu-1-\phi}}$, and $\lambda_3 = \lambda^{\frac{1-\mu}{\mu-1-\phi}}$ are constants.

For constrained firms, Equation (3) is binding, such that the Lagrange multiplier on the cash-flow constraint is positive. This means that the $mrpm^{constrained} > mrpm^{unconstrained} = \frac{1}{\mu}$. If we assume z_j is constant across firms, $mrpm_j$ is decreasing in d_j. Therefore, only firms with sufficiently high *effective demand*, d_j, are operating at their optimum level. Lemma 1 summarises and compares the outcomes for constrained and unconstrained firms.

Lemma 1. *The constrained firms operate at a sub-optimal level. They face a higher shadow cost of variable input. Their marginal revenue product of variable input is higher relative to unconstrained firms, thus, indicating inefficient allocation of variable input. Note that this can be empirically observed as dispersion in $mrpm$, or $Var(mrpm) > 0$.*

Lemma 1 shows that demand-side distortions feed cash-flow based financial constraints to generate dispersion in the marginal product of material input ($mrpm$). Hence, groups of firms that face different composition of demand will operate, in equilibrium, at different $mrpm$. Firms at a higher $mrpm$ are forced to remain small, thus contributing to inefficiencies in the aggregate economy. Therefore, the dispersion in $mrpm$ captures the extent of misallocation in variable input, between different groups of firms in the economy. In our empirical analysis, we will use this measure to investigate the existence of misallocation in rural India, due to working-capital constraints.

3.3 Equilibrium

To understand the composition of constrained and unconstrained firms in rural India, we apply the above framework and derive the relationship between firm size and *effective demand*. To work with a unique interior equilibrium, we begin by assuming that the collateralizability of cash flow for a firm can only be as strong as the size of *effective demand* for the firm. In other words, a product with low substitutability translates into higher *effective demand*, higher collateralizability and a weaker working-capital constraint. We summarise this notion formally in Assumption 1.

Assumption 1. *Feasible values of ϕ are bound by the elasticity of substitution between firms. Therefore, we focus on the parameter space where $1 + \phi < \frac{\sigma}{\sigma-1}$.*

Note that firm revenue is concave in variable input, v_j, with returns to scale given by $\frac{1}{\mu} < 1$. Thus, for every additional unit of variable input, a firm generates less than one unit of revenue. It is possible that with $\phi > 0$, a firm may be able to escape these decreasing

returns parameter-space to a certain extent. For instance, if $1 + \phi = \frac{1}{\mu}$, then depending on the value of constants, either all firms are constrained or all firms are unconstrained. Assumption 1 helps us focus on the parameter-space where revenue-based working-capital constraints are binding only for some fraction of firms, and therefore, both constrained and unconstrained firms co-exist in the economy. To formalise this, Proposition 1 derives a threshold, which summarises the measure of constrained and unconstrained firms in the economy.

Proposition 1. *Under Assumption 1, there exists a unique level of effective demand, $\widehat{d}$, such that, for all firms with $d_j < \widehat{d}$, the borrowing constraint is binding.*

Proposition 1 follows from Equations (2) and (4). The optimal level of variable input in Equation (2) increases linearly with demand, that is $v^* \propto d_j$. However, the actual level of v_j in Equation (4) is increasing and convex in d_j. Therefore, the two curves intersect once, which pins down $\widehat{d}$, as illustrated in Figure 1.

In our empirical analysis, we will consider an exogenous classification of firms based on their owner's caste. Proposition 1 implies that if firms owned by members of historically disadvantaged castes cater to households with low income (low d_j), then they are more likely to be subject to stunted growth. We will pursue this hypothesis further in our empirical analysis.

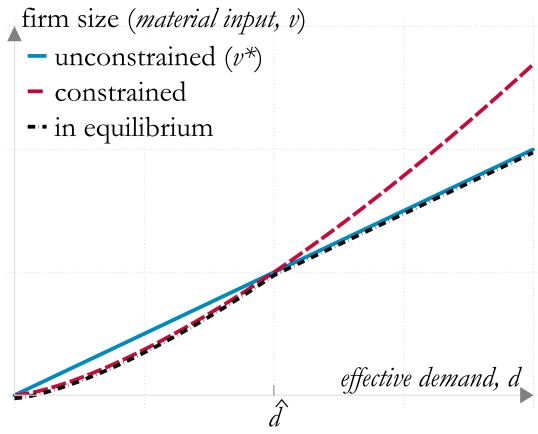
3.4 Comparative Statics

In our model, the wedges induced due to households' preferences and income distribution in a local economy can be summarized by the asymmetry in effective demand, d_j, faced by different firms. Suppose that the income distribution changed in such a way that the effective demand faced by some firms changed from d_j to d'_j. How would it affect the growth of firms? Would it affect all firms uniformly? How would it contribute to misallocation in the local economy, as measured by $mrpm$? Taking ϕ, σ as primitives, the following propositions show how one can compute the impact of such a change.

Proposition 2. *A positive shift in demand increases firm size for all firms, but relatively more for constrained firms.*

The elasticity of revenue to demand shifts for unconstrained firms is one, whereas in the case of constrained firms, the elasticity is $\alpha_1 = \frac{\mu}{\sigma(\mu-1-\phi)}$. Under Assumption 1 and $\phi > 0$, α_1 is strictly greater than 1. This is because a positive demand shift has a direct scale effect on all firms and an additional indirect effect through cash-flow constraints only on constrained firms. This model prediction implies that, in the context of our empirical setting, when there is shift in income (or d_j), firms owned by members of historically disadvantaged castes (LC firms) that were earlier constrained due to low demand, will

Figure 1: Distribution of firm size is characterised by firms' *effective demand*



now grow more than other firms. Further, if the set of households LC firms cater to are affected by a relatively larger income shift than other households, one can expect a further amplification in the growth of LC firms over HC firms. We will estimate the revenue elasticity of different firms to such a local income shift, in our empirical analysis.

While Proposition 2 describes how an income shift would alter outcomes at the firm-level, it is illustrative to study its effect on $mrpm$, to shed light on the change in aggregate inefficiencies in the local economy. Proposition 3 summarises the result of this exercise.

Proposition 3. *A positive demand shift decreases $mrpm$ only among constrained firms, and $mrpm$ remains constant among unconstrained firms, thus reducing $Var(mrpm)$ in the economy.*

As shown in Equation (2), $mrpm$ is independent of d_j for unconstrained firms. But, as shown in Equation (4), for constrained firms, the $mrpm$ is strictly decreasing with d_j at an elasticity of $\frac{-\mu\phi}{\sigma(\mu-1-\phi)}$. A positive demand shift relaxes individual firms' working-capital constraints, which, in aggregate, reduces the mass of constrained firms in the economy (pushing more firms' d_j to the right of $\widehat{d}$). In the context of our empirical setting, when there is shift in income (or d), more number of LC firms will now operate at their optimal size, thus closing the gap in $mrpm$ between LC firms and HC firms and reducing misallocation of *variable input* across firms in the local economy. Further, if the set of households LC firms cater to are affected by a relatively larger income shift than other households, one can expect a further reduction in misallocation. We will estimate the elasticity of $mrpm$ of different firms to such a local income shift, in our empirical analysis.

3.5 Model Discussion

The model builds on the premise that firms in rural India are financially constrained and rely on working-capital for expansion. However, due to historical segregation in

the social structure, firms are exposed to asymmetric demand compositions. The key identification assumption we make in our model is that demand shifters move variable input and labor, but not the use of homogeneous input such as land, machinery, etc.[11] In other words, to document the wedges induced in the economy due to this demand-induced distortion, we assume that a relaxation of firms' constraints affects their *variable input* and *labor* procurement immediately, with no immediate effect on homogeneous inputs such as land, infrastructure, etc.[12]

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 dataset is that it contains micro-level data on the caste of both, firm owners and households, in every district in India. It also contains variables on the consumption behaviour, wages, demographics and firm production.

We begin by documenting the economic conditions of individuals and households historically classified as members of different castes. Next, we document asymmetries in firm-level outcomes across caste.

4.1 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.[13] The descriptive statistics are provided in Table C.1. HC workers account for 33.3% of the employment survey sample, while LC workers account for 28.9% of the sample. However, the HC workers report, on average, a higher daily wage (84 Indian rupees) compared to LC workers (48 Indian rupees), and a lower employment in agriculture (28% for HC compared to 42% for LC).

[11] See, for instance, Boehm and Oberfield (2020) for a previous application of the identification assumption.

[12] See, for instance, Cooper and Haltiwanger (2006) for the importance of non-convexities in capital adjustment costs.

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

4.2 Consumption Data

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

Panel A and B in Table C.1 together suggest that LC consumption is constrained and much lower than HC consumption. This suggests that the *effective demand* faced by firms which cater to a larger fraction of LC population are more likely to fall below the working-capital-constraint threshold, $\widehat{d}$ (see Proposition 1). This implies that LC firms are more often bound by working-capital constraints than HC firms, and may not achieve their potential size. To find evidence on firms' outcomes, we turn to the MSME dataset.

4.3 Micro Small and Medium Enterprise (MSME) Data

The MSME dataset consists of two parts: a census of registered MSMEs and a survey of unregistered MSMEs.[14] In particular, the dataset 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 value added* as the measure of revenue or output. The amount of loan outstanding captures all the loans from formal and informal

[14] 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."

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. The descriptive statistics are provided in Table C.1.

Panel C in Table C.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 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 total gross output value in the data, while HC firms account for 61.3%. In terms of material input value, LC firms account for 5.4% of 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%. However, Panel C in Table C.1 also shows that firms owned by members of historically disadvantaged castes (LC firms) are more productive than others. When productivity is measured in terms of output-labor ratio or in terms of output-material input ratio, LC firms show a mean ratio of 1.30 compared to 1.21 for HC firms. The above evidence suggests that firm sizes in India are distorted and resources may be inefficiently allocated, across caste. Figure C.1 plots the spatial variation in this asymmetry in entrpreneurial 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.

MSME Retrospective Panel 2005-2007: The MSME census also provides retrospective information on *gross value added* for the firms that survive up to 2007. This allows us to construct a balanced panel of MSMEs for the three-year period, 2005-2007. This synthetic panel allows us to compute statistics such as the auto correlation of firms' output (see Appendix B for more details).

Cross-caste patterns from Tables Table C.1 are consistent with the model prediction from Proposition 1 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, we will combine these cross-caste patters with asymmetric cross-caste effects of exogenous deviations in monsoon rainfall. 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). The 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.[15] For Indian districts between 1980 and 2016, monsoon rainfall accounts for more than 75% of annual rainfall[16]. Figure C.2 plots the spatial variation in rainfall, measured in millimeters of monsoon rainfall. The plot shows a lot variation across districts in India.

Consider the two sources of cross-sectional variation plotted in Figures C.1 and C.2. 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, we 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

Guided by our model predictions from Section 3, and 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.[17]

Our theoretical predictions build on the premise that historically disadvantaged community, denoted by LC, are poor and constrained (see Table C.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 monsoon rainfall, and observe its effects on firms owned by different caste-members across different sectors.[18] 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.[19]

[15] 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.

[16] 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).

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

[18] See Section 6.2 for alternative measures of rainfall as a demand shifter.

[19] This is a standard assumption in prior work. See, for instance, Jayachandran (2006), Kaur (2019), and Santangelo (2019). In Section 6.2, we also document the lack of serial correlation in monsoon rainfall, to rule out persistence in the occurrence of higher rainfall.

5.1 Estimation

5.1.1 Effect of rainfall on individual wages, across caste

While previous work has already shown the effect of rainfall on average local wages, it is necessary that we frame our setting by first providing evidence on the asymmetry in economic conditions across caste and the asymmetric effects of rainfall on income across caste.[20] 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 monsoon 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 labor force in rural India and have low income-levels, higher monsoon rainfall significantly improves their income.

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

$$log(w_{it}) = \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_{dt} \times caste_i + \gamma X_{it} + \delta_d + \delta_t + \epsilon_{it} \tag{5}$$

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 caste characteristics. Second, we control for any time-invariant district characteristics. The third is time fixed effect. We also include controls at the individual level: age, gender, and crop season.[21] 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.[22] The coefficient $\beta_{3,LC}$ gives the elasticity of LC workers' wages to monsoon rainfall. According to our hypothesis, LC workers have lower income levels at baseline ($\beta_{2,LC} < 0$), but higher monsoon 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 monsoon rainfall especially relax the constraints of LC

[20] 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.

[21] 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.

[22] See Section 6.2 for an alternative specification of standard errors.

households and increase their consumption, through the effect of rainfall on agricultural wages. This helps us establish that monsoon 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 are also able to establish the effect of monsoon rainfall as a demand-shifter across non-agricultural product categories.

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

$$log(c_{ht}) \;=\; \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_{dt} \times caste_i + \delta_d + \delta_t + \epsilon_{it} \quad (6)$$

where h denotes household. The regression includes the two-way fixed effects as in the previous equation. The coefficient $\beta_{3,LC}$ gives the elasticity of LC households' consumption to monsoon rainfall. According to our hypothesis, LC households have lower consumption levels at baseline ($\beta_{2,LC} < 0$), but higher monsoon 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 us establish the importance of caste-linkages in determining firms' demand in the local rural economy.

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

$$log(y_{ft}) \;=\; \alpha + \beta_1 \cdot rainfall_{dt} + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_{dt} \times caste_i + \delta_d \times \delta_t + \delta_F + \epsilon_i \quad (7)$$

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 × year fixed effect to control for any time-varying district-specific characteristics. 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 monsoon rainfall.

If a certain group of firms' demand is largely composed of constrained households', then Proposition 1 predicts that these firms are more likely to exhibit stunted growth. We hypothesis that LC firms' demand is linked to LC households' consumption, and therefore,

are more likely to be smaller. According to our hypothesis, we expect $\beta_{2,LC} < 0$ in Columns (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, guided by Proposition 2, we expect $\beta_{3,LC} > 0$ in Columns (1) and (2). According to Proposition 2, an exogenous shift in LC households' demand translates into higher revenue for LC firms, by increasing their cash-flow and relaxing their working-capital constraints, thus allowing them to grow closer to their optimal size.

Stunted growth of some firms is inefficient for the aggregate economy, especially if they are highly productive enterprises. According to Proposition 3, if LC firms are more productive but constrained from reaching their optimal size, we should firstly observe a dispersion in productivity across caste, as measured by the Marginal Revenue Product of Materials ($mrpm$). Secondly, as constraints on LC firms relax, due to the shift LC demand following higher monsoon rainfall, and they perform closer to their potential, we should observe a reduction in the productivity dispersion in the economy. Therefore, we expect $\beta_{2,LC} < 0$ and $\beta_{3,LC} > 0$ in Column (3). This can be interpreted as an indication of the inefficiencies in allocation of *variable input* across firms owned by members of different castes.

5.1.4 Mechanism

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 monsoon rainfall on firms owned by members across different caste, we estimate the following equation:

$$log(y_f) \;=\; \alpha + \beta_1 \cdot rainfall_d + \beta_{2,i} \cdot caste_i + \beta_{3,i} \cdot rainfall_d \times caste_i + \delta_d + \delta_F + \epsilon_i \qquad (8)$$

First, we estimate the effect of higher rainfall on the following firm-level variables (y): (1) revenue, (2) material input, (3) wages paid, (4) ratio of revenue to material input, and (5) ratio of material input to wages paid. This exercise sheds light on two issues: (i) to check that our results from the previous panel regression hold in the cross-section as well, and (ii) to provide evidence on the nature and intensity of constraint that firms owned by different caste-members face. In particular, we compare the constraints faced by the firms in procuring variable input to paying wages.

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 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 6.2.

6 Results

Individual wage elasticity: Table 1 presents the results from estimating Equation (5), and shows the effect of higher rainfall on local wages. First, it shows the asymmetry in economic conditions across caste. LC agricultural workers earn 4.38% lower wages than HC agricultural workers, in a district with median rainfall. Second, wages of LC agricultural workers increase by 7% 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 2 presents the results from estimating Equation (6), and shows the effect of higher rainfall on Monthly Per Capita Expenditure (MPCE). First, it 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 3 presents the results from estimating Equation (7), and shows the effect of higher rainfall on firms' equilibrium outcomes. First, it shows that LC firms are 62.8% smaller (by revenue) than HC firms, in a district with median

rainfall. This is consistent with the prediction from Proposition 1 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 2, 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 Proposition 2, 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.

Aggregate inefficiency: Third, Column (3) in Table 2 shows that $mrpm$ for LC firms is higher than HC firms by 16.1%, in a district with median rainfall, suggesting that LC firms' constraints are limiting the amount of material input they are able to procure for production. This can be explained by both financial market frictions or a large share of LC firms' demand being composed of constrained LC consumers.

However, Table 2 also shows that, consistent with Proposition 3, due to higher rainfall, a rise in LC demand reduces the working-capital constraints on LC firms which allows them to procure more material inputs and expand their production. As these productive firms are able to grow, this reduces the dispsersion in productivity in the local economy, as measured by $mrpm$ (5.3% decrease). This result suggests that when demand is determined by caste-linkages, firms reflect the constraints of their consumers, and once those constraints are relaxed, the firms are also able to grow closer to their optimal size.

6.1 Additional evidence

Finally, we investigate the nature of growth shown by LC firms and the mechanism through which higher rainfall relaxes the constraints of an LC firm. In order to do so, we focus on the cross-section of MSME firms in rural India in 2006 - 07.

Labor: In Table 4, we begin by establishing that the results from the MSME panel shown in Table 3 hold in the cross-section as well. Secondly, Table 4 shows that wages paid by LC also increase by 13.1%, 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 5 shows that there is no significant difference in the prices of LC firms relative to HC. 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 6 shows that value of loans LC firms have taken is 68.5% lower than that of HC firms, in a district with median rainfall.[23] 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. This evidence provides further support in the identification of cash-flow based constraints among firms in rural India.

6.2 Robustness Checks

Table C.2 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, where the sample was restricted to workers between the ages of 18 and 60.

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 C.3 are consistent with the baseline results from Table 2. 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.

In Table C.4, 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

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

patterns documented in Table 2 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 2.

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 C.5, by controlling for land-owned, education, meals consumed by the head of household at the district×caste×year level, interacted with monsoon. We find that firms' outcomes are qualitatively similar to the baseline Table 2. Tables C.6 - C.8 show that the results in Tables 4 - 6 are also robust to this concern.

In Tables C.9 - C.11, we cluster standard errors at the state-year level to allow for spatial correlation. The results are qualitatively consistent with our baseline results from Tables 1 - 3.

In Tables C.12 - C.14, we use an alternative definition of monsoon. 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 Tables 1 - 3.

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

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 communities (denoted by LC) earn lower wages and that their consumption is constrained, relative to others. Higher rainfall improves crop productivity and the wages of the LC workers 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 dataset and deserves further research.

Main Tables

Table 1: Wage Elasticity in Rural India: NSS 2004-10

	(1) All	(2) Agri.	(3) Mfg.
log(rainfall)	0.043**	0.037*	0.009
	(0.019)	(0.020)	(0.074)
MC	-0.183	-0.207*	-0.134
	(0.159)	(0.120)	(0.373)
LC	-0.374***	-0.250**	-0.113
	(0.141)	(0.103)	(0.377)
log(rainfall) × MC	-0.003	0.028	-0.006
	(0.024)	(0.018)	(0.054)
log(rainfall) × LC	0.012	0.033**	-0.019
	(0.021)	(0.016)	(0.055)
Observations	129,988	61,571	10,296
R-squared	0.321	0.351	0.470
District FE	✓	✓	✓
Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: The regressions of individual-level daily wage (logarithmic) on monsoon. $Agri.$ stands for agricultural workers in rural areas. $Mfg.$ stands for manufacturing workers in rural areas. The additional control variables are age, gender, and crop season. Sample includes individuals between the ages of 18 and 60. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, ***p<0.01, **p<0.05, *p<0.1.

Table 2: Household Consumption Elasticity in Rural India (by caste): NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
log(monsoon)	-0.005	-0.044	-0.002	-0.095
	(0.021)	(0.028)	(0.048)	(0.061)
MC	-0.583***	-0.440***	-1.357***	-0.951**
	(0.147)	(0.138)	(0.380)	(0.415)
LC	-1.015***	-0.677***	-1.706***	-2.326***
	(0.138)	(0.162)	(0.343)	(0.377)
log(monsoon) × MC	0.058***	0.042**	0.155***	0.113*
	(0.022)	(0.020)	(0.056)	(0.062)
log(monsoon) × LC	0.100***	0.048**	0.138***	0.280***
	(0.021)	(0.024)	(0.051)	(0.056)
Observations	154,241	153,586	132,873	126,125
R-squared	0.351	0.251	0.225	0.218
District FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: The regressions of household' monthly per capita consumption (logarithmic) on monsoon. $Mfg.$ 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.

Table 3: Firm-level Elasticities in Rural India: MSME 2004-07

	(1)	(2)	(3)
	Revenue	Material Input	$\frac{\text{Revenue}}{\text{Material Input}}$
MC	-1.443***	-1.750***	0.318*
	(0.308)	(0.417)	(0.183)
LC	-1.763***	-2.273***	0.511***
	(0.314)	(0.400)	(0.144)
log(monsoon) × MC	0.158***	0.191***	-0.035
	(0.045)	(0.061)	(0.027)
log(monsoon) × LC	0.172***	0.225***	-0.053**
	(0.046)	(0.059)	(0.021)
Observations	1,468,689	1,457,160	1,451,384
R-squared	0.422	0.460	0.329
District × Year FE	✓	✓	✓
Product FE	✓	✓	✓

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

Table 4: Firm-level Elasticities in Rural India: MSME 2007

VARIABLES	(1) Revenue	(2) Material Input	(3) Wages	(4) $\frac{\text{Revenue}}{\text{Material Input}}$	(5) $\frac{\text{Material Input}}{\text{Wages}}$
MC	-1.283***	-1.595***	-0.877***	0.316**	-0.592***
	(0.311)	(0.386)	(0.229)	(0.159)	(0.220)
LC	-1.497***	-1.973***	-0.897***	0.484***	-0.905***
	(0.315)	(0.385)	(0.232)	(0.139)	(0.220)
log(monsoon) × MC	0.133***	0.167***	0.096***	-0.035	0.054*
	(0.045)	(0.056)	(0.034)	(0.023)	(0.032)
log(monsoon) × LC	0.131***	0.179***	0.071**	-0.049**	0.086***
	(0.046)	(0.057)	(0.034)	(0.021)	(0.033)
Observations	511,137	508,871	483,151	508,857	481,269
R-squared	0.423	0.458	0.469	0.326	0.334
District FE	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓

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

Table 5: Output Price and Input Price Elasticity in Rural India: MSME 2007

	(1) Output Price	(2) Input Price
MC	-0.594*	-0.855*
	(0.346)	(0.467)
LC	-0.321	0.089
	(0.352)	(0.435)
log(monsoon) × MC	0.071	0.103
	(0.050)	(0.066)
log(monsoon) × LC	0.027	-0.063
	(0.052)	(0.063)
Observations	267,912	168,946
R-squared	0.358	0.393
District FE	✓	✓
Product FE	✓	✓

Notes: The regressions of product-level prices on monsoon in the year 2007. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Firm-level Loans Taken in Rural India: MSME 2007

	(1) All	(2) Institutional	(3) Non-institutional
MC	-0.606	-0.612	0.266
	(0.488)	(0.445)	(1.200)
LC	-1.404*	-0.965	-0.141
	(0.716)	(0.630)	(2.493)
log(monsoon) × MC	0.023	0.031	-0.093
	(0.068)	(0.064)	(0.174)
log(monsoon) × LC	0.109	0.055	-0.132
	(0.105)	(0.092)	(0.361)
Observations	33,541	28,666	1,404
R-squared	0.425	0.482	0.678
District FE	✓	✓	✓
Product FE	✓	✓	✓

Notes: The regressions of firm-level loans taken on monsoon in the year 2007. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** p<0.01, ** p<0.05, * p<0.1.

References

[Citing pages are listed after each reference.]

Banerjee, Abhijit, and Kaivan Munshi. How efficiently is capital allocated? evidence from the knitted garment industry in tirupur. *The Review of Economic Studies*, 71(1):19–42, 2004. [see page: 3]

Bernard, Andrew B, Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes. The origins of firm heterogeneity: A production network approach. *Journal of Political Economy*, 130(7):1765–1804, 2022. [see page: 4]

Bidner, Chris, and Mukesh Eswaran. A gender-based theory of the origin of the caste system of india. *Journal of Development Economics*, 114:142–158, 2015. [see page: 4]

Boehm, Johannes, and Ezra Oberfield. Misallocation in the market for inputs: Enforcement and the organization of production. *The Quarterly Journal of Economics*, 135(4):2007–2058, 2020. [see pages: 1, 11]

Cassan, Guilhem, Daniel Keniston, and Tatjana Kleineberg. A division of laborers: Identity and efficiency in india. Technical report, National Bureau of Economic Research, 2021. [see page: 3]

Cooper, Russell W, and John C Haltiwanger. On the nature of capital adjustment costs. *The Review of Economic Studies*, 73(3):611–633, 2006. [see page: 11]

Deschênes, Olivier, and Michael Greenstone. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American economic review*, 97(1): 354–385, 2007. [see page: 14]

Deshpande, Manali S. History of the indian caste system and its impact on india today. 2010. [see page: 4]

Drechsel, Thomas. Earnings-based borrowing constraints and macroeconomic fluctuations. *DP16975*, 2022. [see pages: 1, 5]

Einav, Liran, Peter J Klenow, Jonathan D Levin, and Raviv Murciano-Goroff. Customers and retail growth. Technical report, National Bureau of Economic Research, 2021. [see page: 3]

Foster, Lucia, John Haltiwanger, and Chad Syverson. The slow growth of new plants: Learning about demand? *Economica*, 83(329):91–129, 2016. [see page: 3]

Goraya, Sampreet Singh. How does caste affect entrepreneurship? birth versus worth. *Journal of Monetary Economics*, 2023. [see pages: 3, 5, and 20]

Hjort, Jonas. Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4): 1899–1946, 2014. [see page: 3]

Hottman, Colin J, Stephen J Redding, and David E Weinstein. Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364, 2016. [see page: 3]

Hsieh, Chang-Tai, and Peter J Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448, 2009. [see page: 7]

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow. The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474, 2019. [see page: 3]

Jayachandran, Seema. Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of political Economy*, 114(3):538–575, 2006. [see pages: 2, 14, 15, and 21]

Jensen, Robert, and Nolan H Miller. Market integration, demand, and the growth of firms: Evidence from a natural experiment in india. *American Economic Review*, 108(12):3583–3625, 2018. [see page: 1]

Kaur, Supreet. Nominal wage rigidity in village labor markets. *American Economic Review*, 109(10):

3585–3616, 2019. [see pages: 2, 14, and 15]

Lian, Chen, and Yueran Ma. Anatomy of corporate borrowing constraints. *The Quarterly Journal of Economics*, 136(1):229–291, 2021. [see page: 5]

MSME. *Fourth All India Census of Micro, Small and Medium Enterprises, 2006*. Ministry of Micro, Small and Medium Enterprises, 2009. URL https://www.dcmsme.gov.in/ito_msme/censuses.htm. [see page: 32]

Munshi, Kaivan. Caste and the indian economy. *Journal of Economic Literature*, 57(4):781–834, 2019. [see pages: 1, 4]

Munshi, Kaivan, and Mark Rosenzweig. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(01):46–98, 2016. [see page: 3]

Oh, Suanna. Does identity affect labor supply? *Available at SSRN 3998025*, 2021. [see page: 3]

Santangelo, G. Firms and farms: The local effects of farm income on firms' demand. *Cambridge Working Papers in Economics*, 2019. [see pages: 2, 14, 14, 14, and 15]

A Model: Proofs

Lemma 1. Unconstrained Problem

$$\max_{v} \pi_j(z,v) = p_j y_j - v_j \tag{9}$$

$$s.t. \quad p_j = d_j^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}} \tag{10}$$

First-order conditions w.r.t. variable input is

$$\left(\frac{\sigma-1}{\sigma}\right) \underbrace{\frac{p_j y_j}{v_j}}_{mrpv_j} = 1 \implies v_j = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} d_j z_j^{\sigma-1} \tag{11}$$

One can write revenue as

$$r_j = p_j y_j = d_j^{\frac{1}{\sigma}} y_j^{\frac{\sigma-1}{\sigma}} = \left(\frac{\sigma-1}{\sigma}\right)^{\frac{\sigma}{\mu}} d_j z_j^{\frac{\sigma}{\mu}} \tag{12}$$

The firm problem with borrowing constraints is

$$\max_{v} \pi_j(z,v) = p_j y_j - v_j \tag{13}$$

$$s.t. \quad p_j = d_j^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}} \tag{14}$$

$$v_j \geq \lambda r_j^{1+\phi} \tag{15}$$

The constrained optimal gives us the value of variable inputs

$$v_j = \lambda r_j^{1+\phi} = \lambda d_j^{\frac{1+\phi}{\sigma}} z_j^{\frac{(1+\phi)(\sigma-1)}{\sigma}} v_j^{\frac{(1+\phi)(\sigma-1)}{\sigma}} \implies v_j = \underbrace{\lambda^{\frac{\mu}{\mu-1-\phi}}}_{\lambda_1} d_j^{\frac{\mu(1+\phi)}{\sigma(\mu-1-\phi)}} z_j^{\frac{1+\phi}{\mu-1-\phi}} \tag{16}$$

and revenue for the constrained firms is

$$r_j = d_j^{\frac{1}{\sigma}} z_j^{\frac{\sigma-1}{\sigma}} v_j^{\frac{\sigma-1}{\sigma}} = \underbrace{\lambda^{\frac{1}{\mu-1-\phi}}}_{\lambda_2} d_j^{\frac{\mu}{\sigma(\mu-1-\phi)}} z_j^{\frac{1}{\sigma(\mu-1-\phi)}} \tag{17}$$

The margin product of variable inputs for constrained firms is

$$mrpv_j = \frac{r_j}{v_j} = \frac{\lambda_2}{\lambda_1} d_j^{\frac{-\mu\phi}{\sigma(\mu-1-\phi)}} z_j^{\frac{-\phi}{\sigma(\mu-1-\phi)}} \tag{18}$$

Proposition 1. For any firm j, the demand for variable input is constrained if

$$v_j \leq v^* \tag{19}$$

$$\lambda_1 \, d_j^{\frac{\mu(1+\phi)}{\sigma(\mu-1-\phi)}} \, z_j^{\frac{1+\phi}{\mu-1-\phi}} \leq \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} d_j z_j^{\sigma-1} \tag{20}$$

Let us define the threshold level of demand $\widehat{d}$ such that the equation 20 hold with equality. This gives us

$$\widehat{d} = \lambda^{\frac{-\sigma(\mu-1-\phi)}{\mu(1+\phi)-\sigma\mu+\sigma(1+\phi)}} \, \mu^{\frac{-\mu-1-\phi}{\mu(1+\phi)-\sigma\mu+\sigma(1+\phi)}} \, z^{\frac{(\sigma-1)(\mu-1-\phi)-(1+\phi)}{\mu(1+\phi)-\sigma\mu+\sigma(1+\phi)}} \tag{21}$$

Proposition 2. Using equation 12 and taking logs both sides, we get

$$\log r^* = \frac{\sigma}{\mu} \log\left(\frac{\sigma-1}{\sigma}\right) + \log d_j + \frac{\sigma}{\mu} \log z_j \tag{22}$$

$$\frac{\partial \log r^*}{\partial \log d_j} = 1 \tag{23}$$

Using equation 17 and taking logs both sides, we get

$$\log r^j = log\lambda_2 + \frac{\mu}{\sigma(\mu-1-\phi)} \log d_j + \frac{1}{\sigma(\mu-1-\phi)} \log z_j \tag{24}$$

$$\frac{\partial \log r_j}{\partial \log d_j} = \frac{\mu}{\sigma(\mu-1-\phi)} \tag{25}$$

Proposition 3. For unconstrained firms, we can define

$$\log mrpv^* = \log \frac{1}{\mu} \implies \frac{\partial \log mrpv^*}{\partial \log d_j} = 0 \tag{26}$$

For constrained firms, we can define

$$\log mrpv_j = \log \frac{\lambda_2}{\lambda_1} - \frac{\mu\phi}{\sigma(\mu-1-\phi)} \log d_j + \frac{-\phi}{\sigma(\mu-1-\phi)} \log z_j \tag{27}$$

$$\frac{\partial \log mrpv_j}{\partial \log d_j} = \frac{-\mu\phi}{\sigma(\mu-1-\phi)} \tag{28}$$

B Data: MSME Census

The main data source is the Micro, Small and Medium Enterprises (MSME) data from 2006 to 2007 (MSME, 2009), the details of which are provided below. The main advantage of using the MSME and EC is that they provide the caste of the enterprise owner together with other financial variables. However, MSME is not as commonly used as the ASI-NSS or the CMIE Prowess because it omits large enterprises (above a certain threshold of capital stock), nor does it have a panel dimension.

MSME Dataset: The MSME dataset contains a representative sample of the *MSME sector* that comprises enterprises that invest less than INR 100 million (manufacturing enterprises) and INR 50 million (service enterprises).[24] An enterprise is a firm. The MSME dataset 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. Finally, 1.45 million observations (1.12 million in the manufacturing sector) are left after the cleaning process. The descriptive statistics for manufacturing are provided in Table C.1. 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 C of Table C.1. LC and MC enterprises are smaller and have lower leverage (defined as the ratio of credit to capital stock) than HC enterprises. Next, the retrospective information on output is exploited from the dataset to compute growth rates. The average growth rate of output is around 9% and does not vary a lot across castes, whereas the dispersion in output growth is higher for HC enterprises relative to LC and MC enterprises. Further, the default rates for each caste are computed. Around 14.5% of LC enterprises and 13.6% of HC enterprises defaulted in the last 12 months.[25]

The MSME dataset 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 that are below a certain threshold of size (size in terms of original value of investment in plant, machinery and equipment). The investment limit for enterprises engaged in the manufacturing or production of goods is Indian rupees (INR) 100 million whereas for those providing or rendering in services is INR 50 million.

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

[24] The MSME sector is estimated to employ about 59 million individuals in over 26.1 million enterprises throughout the country. Further, 1.5 million (5.94%) are registered MSMEs and 24.5 million (94.06 %) are unregistered MSMEs. The registration is under the Factories Act of 1948.

[25] These numbers suggest that, at least among MSMEs, riskiness may not be the primary driver of low credit allocation to LC entrepreneurs.

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 dataset 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.

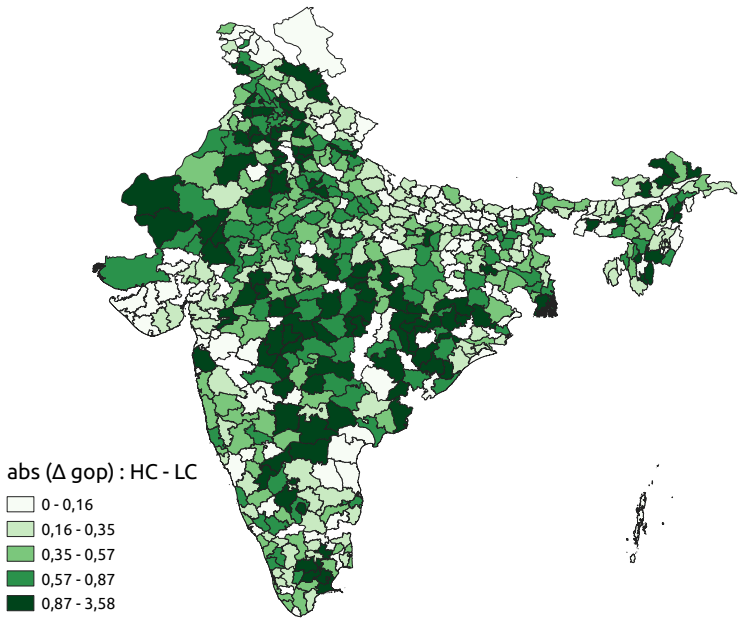
In particular, the MSME dataset provides geographical information, industry classification, balance sheet variables, and the caste of the owner and employees. The measure of capital stock in the data is the market value of fixed assets. The total wage bill includes salaries and wages, allowances, bonuses, and so on. The measure of output is gross value added. The amount of loan outstanding captures all the loans from formal and informal sources, where informal sources include local moneylenders, friends and relatives.

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 inputs and net-worth are winsorized at 1 and 99th percentile. Furthermore, the variables used in regressions $arpk$, $arpl$, k/l are winsorized at 1 and 99th percentile.

MSME Panel 2005-2007: The MSME datset also provides retrospective information on output for the enterprises that survive upto 2007. This allows me to construct a balanced panel of MSMEs for the three-year period, 2005-2007. This panel information allows me to compute statistics such as average output growth, and volatility of output growth for each caste.

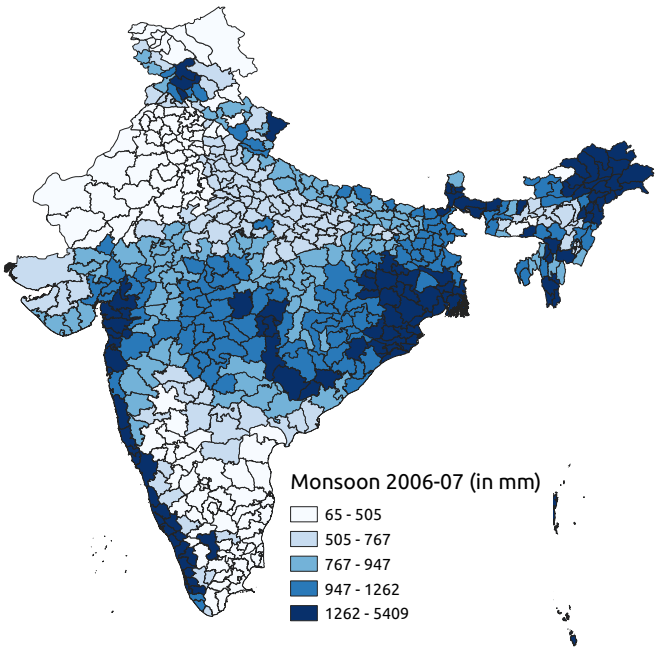
C Figures and Tables

Figure C.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.

Figure C.2: 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.

Table C.1: Summary Statistics

	Panel A: NSS 2004-2010 - Individual-level statistics						
	HC		MC		LC		All
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean (S.D.)
Wages	83.74	(112.68)	54.91	(66.75)	48.3	(56.1)	58.35 (74.32)
Sex (male)	0.79	(0.40)	0.70	(0.46)	0.69	(0.46)	0.71 (0.45)
Age	35.06	(12.32)	35.11	(12.63)	34.62	(12.68)	34.89 (12.60)
Employed in agri.	0.28	(0.45)	0.38	(0.49)	0.42	(0.49)	0.37 (0.48)

	Panel B: NSS 2004-2008 - Household-level statistics						
	HC		MC		LC		All
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)	Mean (S.D.)
MPCE	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)

	Panel C: MSME 2006-2007 - 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)
log(Output/Labor)	1.21	(0.89)	1.06	(0.79)	1.30	(0.82)	1.16 (0.84)
log(Output/Materials)	1.55	(1.05)	1.62	(1.01)	1.77	(1.04)	1.62 (1.04)

Notes: The table reports descriptive statistics. LC, MC, and HC represent entrepreneurs historically classified as belonging to the low-caste, middle-caste, and high-caste, respectively.
Panel A reports statistics for the NSS Employment and Unemployment dataset. 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. The percentage column reports respective caste shares in the full sample value. Wages is measured as the daily wage of workers, deflated for inflation, in Indian rupees. Sampling multipliers are applied.
Panel B reports household-level statistics for the NSS Household Consumption Expenditure dataset. S.D. is the standard deviation. Mean refers to the mean value. The percentage column reports respective caste shares in the full sample 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. Sampling multipliers are applied.
Panel C reports statistics for the MSME dataset. 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 value-added, Labor is total wages, and Materials is material input value. Sampling multipliers are applied. Sampling multipliers are applied.

Table C.2: Wage Elasticity in Rural India: NSS 2004-10

	(1) All	(2) Agri.	(3) Mfg.
log(rainfall)	0.047**	0.045**	0.014
	(0.020)	(0.021)	(0.075)
MC	-0.148	-0.182	-0.021
	(0.149)	(0.121)	(0.353)
LC	-0.328**	-0.222**	-0.010
	(0.135)	(0.107)	(0.354)
log(rainfall) × MC	-0.007	0.023	-0.020
	(0.022)	(0.018)	(0.051)
log(rainfall) × LC	0.008	0.028*	-0.033
	(0.020)	(0.016)	(0.052)
Observations	147,187	72,059	11,914
R-squared	0.314	0.343	0.450
District FE	✓	✓	✓
Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: The regressions of individual level wage (logarithmic) on monsoon shock. $Agri.$ stands for agricultural workers in rural areas. $Mfg.$ stands for manufacturing workers in rural areas. The additional control variables are age, gender, 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 C.3: Household Consumption Elasticity in Rural India (by caste): NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
log(monsoon)	-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(monsoon) × MC	0.105**	0.041	0.249**	0.130
	(0.043)	(0.041)	(0.116)	(0.117)
log(monsoon) × 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(monsoon) × 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(monsoon) × MC × land owned	-0.012	-0.004	-0.026	-0.015
	(0.008)	(0.008)	(0.020)	(0.020)
log(monsoon) × 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 household' monthly per capita consumption (logarithmic) on monsoon. $Mfg.$ 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 C.4: Household Consumption Elasticity in Rural India (by caste): NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
log(monsoon)	-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(monsoon) × MC	0.063***	0.052***	0.173***	0.112*
	(0.020)	(0.020)	(0.054)	(0.061)
log(monsoon) × 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(monsoon) × 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(monsoon) × # 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 monsoon. $Mfg.$ 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.

Table C.5: Firm-level Elasticities in Rural India: MSME 2004-07

	(1) Revenue	(2) Material Input	(3) $\frac{\text{Revenue}}{\text{Material Input}}$
MC	-1.650***	-2.057***	0.416***
	(0.303)	(0.380)	(0.147)
LC	-1.890***	-2.473***	0.578***
	(0.314)	(0.394)	(0.134)
log(monsoon) × MC	0.188***	0.235***	-0.049**
	(0.044)	(0.056)	(0.022)
log(monsoon) × LC	0.191***	0.255***	-0.063***
	(0.046)	(0.058)	(0.020)
land owned	0.436	1.022	-0.617
	(0.385)	(0.629)	(0.399)
log(monsoon) × land owned	-0.064	-0.156	0.097
	(0.060)	(0.099)	(0.063)
education (head of HH)	-1.294***	-2.091***	0.762*
	(0.392)	(0.641)	(0.408)
log(monsoon) × education (head of HH)	0.200***	0.327***	-0.122**
	(0.059)	(0.096)	(0.061)
# of meals (head of HH)	0.097	0.053	0.131
	(0.632)	(0.757)	(0.295)
log(monsoon) × # of meals (head of HH)	-0.029	-0.028	-0.014
	(0.092)	(0.113)	(0.046)
Observations	1,468,686	1,457,157	1,451,381
R-squared	0.422	0.461	0.329
District × Year FE	✓	✓	✓
Product FE	✓	✓	✓

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

Table C.6: Firm-level Elasticities in Rural India: MSME 2007

VARIABLES	(1) Revenue	(2) Material Input	(3) Wages	(4) $\frac{\text{Revenue}}{\text{Material Input}}$	(5) $\frac{\text{Material Input}}{\text{Wages}}$
MC	-1.560***	-1.963***	-1.045***	0.408***	-0.817***
	(0.311)	(0.372)	(0.239)	(0.143)	(0.199)
LC	-1.636***	-2.204***	-0.978***	0.579***	-1.081***
	(0.323)	(0.393)	(0.253)	(0.139)	(0.212)
log(monsoon) × MC	0.172***	0.219***	0.120***	-0.048**	0.085***
	(0.045)	(0.054)	(0.035)	(0.021)	(0.029)
log(monsoon) × LC	0.149***	0.210***	0.081**	-0.063***	0.109***
	(0.047)	(0.057)	(0.037)	(0.021)	(0.031)
land owned	4.257***	5.726***	1.986**	-1.473**	3.720***
	(1.214)	(1.535)	(0.838)	(0.595)	(1.010)
log(monsoon) × land owned	-0.632***	-0.857***	-0.304**	0.226**	-0.552***
	(0.189)	(0.241)	(0.129)	(0.092)	(0.157)
education (head of HH)	-3.047***	-4.206***	-1.747***	1.194**	-2.476***
	(0.790)	(1.086)	(0.584)	(0.601)	(0.728)
log(monsoon) × education (head of HH)	0.455***	0.630***	0.268***	-0.181**	0.364***
	(0.117)	(0.161)	(0.085)	(0.086)	(0.108)
# of meals (head of HH)	-2.174**	-2.494*	-0.967	0.276	-1.424
	(1.015)	(1.286)	(0.688)	(0.450)	(0.909)
log(monsoon) × # of meals (head of HH)	0.335**	0.391*	0.150	-0.050	0.228
	(0.161)	(0.205)	(0.107)	(0.069)	(0.141)
Observations	510,166	507,901	482,203	507,887	480,322
R-squared	0.424	0.458	0.469	0.326	0.334
District FE	✓	✓	✓	✓	✓
Product FE	✓	✓	✓	✓	✓

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

Table C.7: Output Price and Input Price Elasticity in Rural India: MSME 2007

	(1) Output Price	(2) Input Price
MC	-0.867**	-1.236***
	(0.347)	(0.424)
LC	-0.588	-0.409
	(0.370)	(0.444)
log(monsoon) × MC	0.109**	0.155**
	(0.050)	(0.061)
log(monsoon) × LC	0.063	0.000
	(0.055)	(0.065)
land owned	2.875*	7.056***
	(1.675)	(2.288)
log(monsoon) × land owned	-0.415	-1.059***
	(0.266)	(0.346)
education (head of HH)	-3.439**	-6.366***
	(1.356)	(1.830)
log(monsoon) × education (head of HH)	0.494**	0.887***
	(0.192)	(0.256)
# of meals (head of HH)	0.730	0.599
	(1.959)	(1.607)
log(monsoon) × # of meals (head of HH)	-0.086	0.056
	(0.295)	(0.251)
Observations	267,258	168,570
R-squared	0.358	0.394
District FE	✓	✓
Product FE	✓	✓

Notes: The regressions of product-level prices on monsoon in the year 2007. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** p<0.01, ** p<0.05, * p<0.1.

Table C.8: Firm-level Loans Taken in Rural India: MSME 2007

	(1) All	(2) Institutional	(3) Non-institutional
MC	-0.768	-0.832*	0.032
	(0.518)	(0.435)	(1.314)
LC	-1.404**	-0.910	2.370
	(0.711)	(0.663)	(3.273)
log(monsoon) × MC	0.049	0.063	-0.066
	(0.072)	(0.063)	(0.190)
log(monsoon) × LC	0.112	0.046	-0.556
	(0.104)	(0.097)	(0.485)
land owned	-0.571	-0.757	-12.145
	(3.858)	(3.045)	(12.324)
log(monsoon) × land owned	0.084	0.107	1.975
	(0.591)	(0.469)	(1.964)
education (head of HH)	-3.368	-4.279*	22.398**
	(2.462)	(2.478)	(10.815)
log(monsoon) × education (head of HH)	0.579*	0.673*	-3.546*
	(0.348)	(0.348)	(1.792)
# of meals (head of HH)	0.088	1.340	-57.933**
	(3.617)	(2.911)	(24.596)
log(monsoon) × # of meals (head of HH)	-0.144	-0.242	9.261**
	(0.534)	(0.433)	(4.057)
Observations	33,517	28,652	1,403
R-squared	0.425	0.483	0.682
District FE	✓	✓	✓
Product FE	✓	✓	✓

Notes: The regressions of firm-level loans taken on monsoon in the year 2007. Sampling multipliers are applied. Standard errors in parentheses are clustered at the district level, *** p<0.01, ** p<0.05, * p<0.1.

Table C.9: Wage Elasticity in Rural India: NSS 2004-10

	(1) All	(2) Agri.	(3) Mfg.
log(rainfall)	0.043**	0.037*	0.009
	(0.022)	(0.022)	(0.056)
MC	-0.183	-0.207	-0.134
	(0.156)	(0.127)	(0.279)
LC	-0.374**	-0.250**	-0.113
	(0.155)	(0.111)	(0.274)
log(rainfall) × MC	-0.003	0.028	-0.006
	(0.023)	(0.019)	(0.040)
log(rainfall) × LC	0.012	0.033**	-0.019
	(0.024)	(0.017)	(0.040)
Observations	129,988	61,571	10,296
R-squared	0.321	0.351	0.470
District FE	✓	✓	✓
Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: The regressions of individual-level daily wage (logarithmic) on monsoon. $Agri.$ stands for agricultural workers in rural areas. $Mfg.$ stands for manufacturing workers in rural areas. The additional control variables are age, gender, and crop season. Sample includes individuals between the ages of 18 and 60. Sampling multipliers are applied. Standard errors in parentheses are clustered at the state-year level, ***p<0.01, **p<0.05, *p<0.1.

Table C.10: Household Consumption Elasticity in Rural India (by caste): NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
log(monsoon)	-0.005	-0.044	-0.002	-0.095
	(0.027)	(0.027)	(0.051)	(0.069)
MC	-0.583**	-0.440***	-1.357***	-0.915
	(0.228)	(0.158)	(0.339)	(0.556)
LC	-1.015***	-0.677**	-1.706***	-2.371***
	(0.266)	(0.268)	(0.533)	(0.620)
log(monsoon) × MC	0.058*	0.042*	0.155***	0.107
	(0.033)	(0.023)	(0.050)	(0.083)
log(monsoon) × LC	0.100**	0.048	0.138*	0.286***
	(0.040)	(0.039)	(0.080)	(0.091)
Observations	154,241	153,586	132,873	126,125
R-squared	0.351	0.251	0.225	0.217
District FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

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

Table C.11: Firm-level Elasticities in Rural India: MSME 2004-07

	(1)	(2)	(3)
	Revenue	Material Input	$\frac{\text{Revenue}}{\text{Material Input}}$
MC	-1.443***	-1.750***	0.318***
	(0.423)	(0.509)	(0.104)
LC	-1.763***	-2.273***	0.511***
	(0.373)	(0.474)	(0.124)
	(0.314)	(0.400)	(0.144)
log(monsoon) × MC	0.158**	0.191**	-0.035**
	(0.061)	(0.074)	(0.015)
log(monsoon) × LC	0.172***	0.225***	-0.053***
	(0.054)	(0.069)	(0.018)
Observations	1,468,689	1,457,160	1,451,384
R-squared	0.422	0.460	0.329
District × Year FE	✓	✓	✓
Product FE	✓	✓	✓

Notes: The regressions of firm-level variables (logarithmic) on monsoon. Sampling multipliers are applied. Standard errors in parentheses are clustered at state-year level, *** p<0.01, ** p<0.05, * p<0.1.

Table C.12: Wage Elasticity in Rural India: NSS 2004-10

	(1) All	(2) Agri.	(3) Mfg.
Mshock	0.001	0.007	-0.036
	(0.016)	(0.015)	(0.056)
MC	-0.201***	-0.025*	-0.170***
	(0.015)	(0.014)	(0.035)
LC	-0.291***	-0.032**	-0.238***
	(0.016)	(0.013)	(0.035)
Mshock × MC	0.027	0.024	0.020
	(0.018)	(0.017)	(0.052)
Mshock × LC	0.040**	0.027*	0.039
	(0.019)	(0.016)	(0.053)
Observations	129,988	61,571	10,296
R-squared	0.322	0.351	0.470
District FE	✓	✓	✓
Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: The regressions of individual-level daily wage (logarithmic) on monsoon dummy. $Agri.$ stands for agricultural workers in rural areas. $Mfg.$ stands for manufacturing workers in rural areas. The additional control variables are age, gender, and crop season. Sample includes individuals between the ages of 18 and 60. Standard errors in parentheses are clustered at the district level, ***p<0.01, **p<0.05, *p<0.1.

Table C.13: Household Consumption Elasticity in Rural India (by caste): NSS 2004-08

	(1) All	(2) Mfg.	(3) Services	(4) Durables
Mshock	-0.006	-0.017	-0.020	-0.100**
	(0.017)	(0.020)	(0.038)	(0.050)
MC	-0.203***	-0.158***	-0.348***	-0.201***
	(0.016)	(0.016)	(0.041)	(0.038)
LC	-0.357***	-0.360***	-0.801***	-0.504***
	(0.016)	(0.019)	(0.039)	(0.040)
Mshock × MC	0.037	0.009	0.103*	0.023
	(0.023)	(0.023)	(0.057)	(0.054)
Mshock × LC	0.059***	0.032	0.088*	0.210***
	(0.022)	(0.026)	(0.052)	(0.059)
Observations	154,241	153,586	132,873	126,125
R-squared	0.349	0.251	0.224	0.216
District FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Notes: The regressions of household' monthly per capita consumption (logarithmic) on monsoon dummy. $Mfg.$ 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.

Table C.14: Firm-level Elasticities in Rural India: MSME 2004-07

	(1)	(2)	(3)
	Revenue	Material Input	$\frac{\text{Revenue}}{\text{Material Input}}$
MC	-0.385***	-0.468***	0.081***
	(0.032)	(0.041)	(0.015)
LC	-0.613***	-0.769***	0.156***
	(0.033)	(0.042)	(0.014)
Mshock × MC	0.007	0.032	-0.028*
	(0.037)	(0.048)	(0.015)
Mshock × LC	0.009	0.062	-0.055***
	(0.047)	(0.061)	(0.020)
Observations	1,468,689	1,457,160	1,451,384
R-squared	0.421	0.459	0.329
District × Year FE	✓	✓	✓
Product FE	✓	✓	✓

Notes: The regressions of firm-level variables (logarithmic) on monsoon dummy. Sampling multipliers are applied. Standard errors in parentheses are clustered at district level, *** p<0.01, ** p<0.05, * p<0.1.

Table C.15: Testing for serial correlation in rainfall

	$\text{MonsDev}_{d,t}$	
	(1)	(2)
$\text{MonsDev}_{d,t-1}$	0.008	-0.015
	(0.034)	(0.041)
$\text{MonsDev}_{d,t-2}$		-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 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.