Demand for Quality, Variable Markups and (Mis)allocation: Evidence from India

Apoorv Gupta
Department of Economics, Dartmouth College *

Abstract

Markups vary systematically across firms and generate misallocation, yet empirical evidence on sources driving markup variation is limited. I study how demand-side factors affect markups. Using detailed firm-product-level data from India, I document that *both* marginal costs and markups are increasing in firm-size. Changes in markups across the firm-size distribution in response to exogenous demand shocks to poor households lend support to the demand-based markup channel: producing better quality and selling to wealthier, less demand elastic households leads larger firms to incur higher costs and charge higher markups. Accounting for the demand-based channel reduces estimated misallocation losses by 30 percent.

Keywords: Productivity, Incomplete Pass-through, Market Power, Income Distribution.
JEL Classification: D24, D61, E20, L1, O47

*Contact: apoorv.gupta@dartmouth.edu.

This version: March, 2024. I am indebted to Effi Benmelech, Nicolas Crouzet, Dean Karlan, Dimitris Papanikolaou and Jacopo Ponticelli for their invaluable inputs, to David Atkin, David Berger, Diego Comin, Adrien Matray, Martí Mestieri, Terry Moon, Nina Pavcnik, Simon Quinn, Chris Snyder, Chad Syverson, Eric Verhoogen, Jonathan Zinman for insightful discussions and their suggestions, and to participants at Chicago Booth, Rochester Simon, Dartmouth, Berkeley Haas, Berkeley ARE, Penn State, Imperial College London, Maryland, Maryland Smith, University of Hong Kong, Bocconi, CAFRAL (Reserve Bank of India), Indian School of Business, Kellogg School of Management, Northwestern University (Macro Lunch, Development Lunch, Applied Micro), PEDL Young Scholars Conference, STEG Annual Workshops, EMCON, and NEUDC for helpful comments. All mistakes are my own. The data used in this paper come from the Ministry of Statistics and Programme Implementation, Government of India. A previous version of this paper was circulated under the title "Firm Heterogeneity, Demand for Quality and Prices: Evidence from India".

1 Introduction

An important line of research in economics documents large differences in productivity across firms, with productivity increasing in firm-size, particularly in developing countries (Hsieh and Klenow 2009; Bartelsman, Haltiwanger, and Scarpetta 2013). Dispersion is also observed in the marginal revenue products of inputs. When viewed through a standard model of demand and production, dispersion in productivity or marginal products suggests the presence of "distortions" that generate misallocation of resources across firms and potentially lowers the aggregate productivity in an economy.[1] More recently, variation in markups across firms is argued to be an important, if not the leading, source driving these distortions (Haltiwanger, Kulick, and Syverson 2018). However, the extent to which markup dispersion is an evidence of allocative inefficiency depends on the underlying sources driving differences in markups.[2] Under the idea that the differences in markups is driven by the presence of exogenous market (i.e., supply-side) frictions, subsidizing high markup firms will improve resource allocation and increase aggregate productivity. If, instead, driven by firms choosing their optimal markups based on heterogeneous demand elasticities (i.e., demand-side factors), potential gains from reallocation will be limited.

In this paper, I seek to provide evidence on the role that demand-side features play in shaping the distribution of markups. Specifically, I explore how differences in consumer preferences in presence of standard firm heterogeneity can allow large systematic differences in markups to persist in equilibrium. Key to the story is the propensity of richer consumers to source higher share of their consumption from firms producing higher quality goods. This segmentation in the output product market, coupled with differences in consumer demand elasticities across the income distribution, generates variation in firms' markups due to differences in their composition of demand. A direct implication is that markup dispersion, while appearing as a sign of resource misallocation under standard consumer preferences, may well be efficient in an economy with income inequality.

I begin by proposing a framework of *demand-based* markup channel. At the heart of this channel is an assortative matching between larger firms and richer consumers on product quality. The approach is motivated by two theoretical ideas. First, firm's physical productivity and input quality are complements in determining output quality; and higher quality is produced by more productive and larger firms, increasing their marginal costs (Kugler and Verhoogen 2011). Second, consumers are asymmetric in income and richer consumers are more willing to pay for product

[1]The measure of productivity considered here is revenue-based productivity. It is the presence of dispersion in revenue productivity, and not *physical* productivity, that is considered evidence of resource misallocation (Foster, Haltiwanger, and Syverson 2008; Eslava, Haltiwanger, and Urdaneta 2023).

[2]The argument that the design of appropriate policy under *endogenous* distortions will need to focus on the specific mechanism generating that distortion is echoed in Rodrik (1987, 904): "[...] different mechanisms will call for different remedies even if they exhibit themselves in an identical divergence between private and social costs. A straightforward matching of policies with the observed market wedges will not work."

quality (Linder 1961). Together, this leads to richer consumers sourcing larger share of their consumption from goods produced by larger firms, particularly in the quality differentiated sector. Because richer consumers are less price sensitive (a fact I document in my data), assortative matching implies that larger firms charge higher markups.

I then document a set of novel stylized facts on the relationship between firms' size, and their marginal costs and markups that are consistent with the demand-based markup channel. I estimate firm-product level markups using the production function approach from the industrial organization literature, in combination with data from Indian Annual Survey of Industries (ASI) — a highly detailed panel on Indian manufacturing firms from 1998 to 2009. An important feature of the data is that it provides information separately on revenues as well as physical units for each firm-product, allowing to calculate product prices (unit values). Dividing the prices by corresponding markups yields marginal costs at the firm-product level. Using these estimates, I first show that marginal costs within a product group are increasing in firm size. Second, and more importantly, I find that the markups are also increasing in firm size. Third, both these relationships are more pronounced in sectors with greater scope for quality differentiation.

The positive association of markups with firm size, while consistent with, does not identify the demand-based markup channel and can be argued to arise from other sources of variable markups.[3] To make progress on this, I propose an empirical research design that uses quasi-exogenous income shocks to poor households, both across regions and over time, as a source of variation in their demand. Because consumers across income levels differ in the shares of their consumption basket sourced from large-, mid-, and small-sized firms, these changes to the demand from poor households' income affect the aggregate demand faced by small- and mid-sized firms more than large firms. I then trace how firms change their markups to changes in their demand (composition). Importantly, I find no affect of demand shocks on firms' marginal costs or its underlying components, including physical productivity and input prices, reassuring that the empirical strategy isolates changes to consumer demand across the income distribution orthogonal to firms' supply curve.

I find that firms lower their markups in response to an increase in poor households' income. The demand-based markup channel above posits that any changes in demand from lower income groups should affect the weighted demand elasticity, and hence markups, only for firms selling to both rich and poor households. These firms are proxied in my data by firms in the middle of the size distribution. I test this hypothesis by examining how firms across the size distribution change their markups. I find a non-monotonic effect on markups across the firm-size distribution in response to

[3] Recent theoretical work on variable markups suggests that larger firms could charge higher markups under variable elasticity of demand arising from sources other than firms' demand composition (Edmond, Midrigan, and Xu 2019), or because they have larger market shares (Atkeson and Burstein 2008).

higher demand from the poor. Specifically, only mid-sized firms lower their markups while they remain unchanged for firms in the lower and upper ranges of the distribution. More importantly, these responses are only present in quality differentiated sectors, where markup variation is expected to be a consequence of differences in demand composition. I argue that this non-monotonic markup response to demand shocks to the poor is unique to the demand-based markup channel, and also provide empirical evidence inconsistent with alternative explanations and mechanisms.

These findings suggest that differences in the slope of demand faced by firms are an important source of markup dispersion in the differentiated sector. What are the consequences for misallocation and aggregate productivity arising from this *demand-based* markup dispersion? Since demand factors are not prone to reallocation, the aggregate productivity gains that could be attained from a reallocation exercise are correspondingly smaller. To quantify the effects of demand factors on misallocation, I consider a policy that subsidizes (taxes) firms with high (low) markups while assuming markup distortion to be exogenous. Specifically, I follow the literature on static misallocation and consider a policy that serves a planner's objective to equalize marginal revenue product (MRP) of inputs across firms within industries under a fixed aggregate supply of resources, while (erroneously) assuming that any MRP variation arises only from exogenous distortions.[4]

The main result from the exercise is that when markups are endogenous, firms adjust their markups in response to tax-subsidy policies. I propose a sufficient statistic — the estimate of pass-through of changes to firm's costs into its prices — for firms' markup adjustment to the reallocation policy, and develop a methodology to separately identify the contribution of supply (firms' market conduct) and demand (firms' slope of demand) factors. The main intuition of the strategy is that firms in homogeneous sector face the same slope of demand, while the slope of demand can vary across firms in differentiated sector due to differences in their demand composition. This intuition — which derives itself from the results presented above — allows to separately identify the supply and demand parameters from estimated markups and pass-through rates.

I estimate firm-level pass-through rates and find them to be decreasing in firm size, with the relationship stronger in differentiated sectors. Because high markup firms also have lower pass-through rates, productivity gains are substantially lower from the reallocation policy. I estimate that the reallocation gains are substantially large (about 47 percent) when pass-through is assumed to be complete. However, once I account for incomplete pass-through across firms estimated from the data, the reallocation gains are 15 percent. This substantial decrease in productivity gains arises

[4]The dispersion in MRP of inputs might be driven by factors such as unobserved heterogeneity in firm productivity, or adjustment costs, or measurement error instead of firm-specific distortions (see the references in the contribution to the literature below). I use dispersion in marginal revenue product of materials which is less susceptible to adjustment costs. Alternatively, I also consider a policy of equating markups across firms. Markup dispersion, unlike dispersion in productivity, is less likely to be driven by unobserved heterogeneity. Finally, I also rule out measurement error as the potential driver of my results by using panel data with a natural experiment.

because of endogenous markup adjustment by firms in response to policies enacted to lower their markups. Part of this markup adjustment arises from high markup firms facing less price-elastic consumers. I show that the contribution of heterogeneous consumer demand is large — reallocation gains are about 33 percent when firms operate in the most competitive environment observed in the data, while holding fixed the estimated demand parameters faced by firms. This implies that the demand-based markup channel lowers productivity gains from reallocation by 30 percent (14 percentage points).

Contribution to the literature. These findings relate to two literatures. First, a recent and important empirical literature shows that markups vary systematically across firm-size distribution (Atkin, Chaudhry, Chaudhry, Khandelwal, and Verhoogen 2015; Faber and Fally 2020; Eslava, Haltiwanger, and Urdaneta 2023). Relative to these papers, I empirically document how firm heterogeneity interacts with differences in consumer preferences and their demand elasticities to generate a positive relationship between markups and firm-size.[5,6] These findings provide empirical support to the models that generate positive correlations between markups and marginal costs with firm size (Baldwin and Harrigan 2011; Johnson 2012).

By assessing the role of variable markups for misallocation losses, this paper relates to Peters (2020); Edmond, Midrigan, and Xu (2019); Haltiwanger, Kulick, and Syverson (2018). Unlike Peters (2020), this paper studies variable markups driven by differences in consumer preferences (i.e. demand). The demand-based source of variable markups is, in spirit, similar to Edmond, Midrigan, and Xu (2019) with few important distinctions. First, while Edmond, Midrigan, and Xu (2019) rely on a representative consumer with Kimball (1995) demand, this paper documents the important role of consumer heterogeneity. Second, while they quantify the aggregate welfare losses from markups, the focus of this paper is primarily on implications of markup dispersion for allocative efficiency. The importance of consumer heterogeneity as a driver of variable markups and incomplete pass-through has been stressed in the literature (e.g, Nakamura and Zerom 2010; Goldberg and Hellerstein 2012; De Loecker and Goldberg 2014). This paper expands on this idea and proposes an analytical methodology that relies on pass-through rate to adjust for bias associated in estimating reallocation gains under variable markups. The advantages are that it is based on

[5] An important work in this area is Faber and Fally (2020) which documents similar patterns on assortative matching in the US. There are few important differences between this paper and theirs. First, unlike their paper, which uses retail scanner data on households consumption basket to estimate *implied* markups with assumptions on market conduct, I use firm-level production data to estimate *true* markup by relying on 'cost-side' approach without imposing assumptions on nature of competition, or consumer demand. Second, I provide an identification strategy to isolate the role of demand composition for markup dispersion. Third, I study the importance of demand-driven markups for allocative efficiency.

[6] Similarly, a large literature in empirical IO has estimated large differences in consumer preferences across specific industries/products (see Gandhi and Nevo (2021) for a detailed review for work in this area). This paper contributes to that literature by assessing how differences in consumer preferences feeds back into the production side in a *systematic* way that generates higher markups for larger firms across multiple industries in the manufacturing sector.

a reduced-form approach rather than structural estimation, and can be applied to data from other settings and countries with relatively minor assumptions on the demand-side or nature of the market structure.

Second, following the seminal work by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), an extensive literature has focused on factors driving dispersion in productivity or MRPs (Baqaee and Farhi 2019b; David and Venkateswaran 2019). These papers, however, treat these firm-specific factors as exogenous distortions. A related set of papers have paid close attention to the correlation between distortions and firm-size (e.g, Bartelsman, Haltiwanger, and Scarpetta 2013). I contribute to this literature by providing an endogenous source behind the dispersion arising from firms' pricing decisions. My findings suggest that productivity dispersion need not just reflect distortions in this literature and may simply be an equilibrium outcome of standard profit maximization problem when the model is correctly specified.

The paper also contributes to the recent literature that uses detailed micro-data to attribute the observed dispersion in productivity or MRP into factors unrelated to misallocation. This includes unobserved heterogeneity in physical productivity (Gollin and Udry 2021), or adjustment costs (Asker, Collard-Wexler, and De Loecker 2014), or model mis-specification (Haltiwanger, Kulick, and Syverson 2018), or measurement error (Bils, Klenow, and Ruane 2018; Rotemberg and White 2020). Relative to this literature, I assess misallocation losses from variable markups and highlight the importance of studying endogenous responses to policies aimed at improving allocative efficiency. Given that pass-through rates have generally been found to be lower than one across multiple settings,[7] failure to adjust for markup responses generates an upward bias in estimated losses from misallocation. This provides a potential explanation for the observation that productivity losses from misallocation reported using the *indirect approach* — that commonly relies on exogenous wedges — are typically larger than the losses reported using the *direct approach* — that relies on studying responses to specific enacted policies (Restuccia and Rogerson 2017).

The rest of the paper is organized as follows. Section 2 provides the empirical framework, describes the methodology to estimate markups and marginal costs from the firm production data, and describes the data. Section 3 presents stylized facts on markups and costs variation across firms, and on variation in price elasticities across consumer income groups. Section 4 describes the empirical strategy to isolate the role of consumer demand behind markup variation, and presents the results. This endogenous markup dispersion matters, in turn, for misallocation losses. Section 5 assess the gains from a reallocation policy under various scenarios, and Section 6 concludes.

[7]For example, Nakamura and Zerom (2010) find a pass-through elasticity of commodity prices of 25 percent in the coffee industry. Similarly, Goldberg and Hellerstein (2012) find a pass-through of only 5 percent of an exchange rate change into final prices of traded goods in the beer industry. More relevant to this paper, using data on US manufacturing firms, Ganapati, Shapiro, and Walker (2020) find an average pass-through of 70 percent and Haltiwanger, Kulick, and Syverson (2018) find an average pass-through rate of 50 percent across 11 industries.

2 Empirical Framework

This section describes my basic framework. I start by deriving a general expression for markups from firms' profit maximization problem. I show that firm markups depends on both demand- and supply-side factors. I then describe the methodology to compute markups and marginal costs from firm-level production data using the firm's dual problem (through its cost minimization) used in the industrial organization (IO) literature. Finally, I relate markup dispersion to resource misallocation.

2.1 An expression for firms' markups under consumer heterogeneity

I start with a simple model of firm production in presence of consumer heterogeneity. I make few simplifying assumptions that are either standard in the literature or are verified through the data in Section 3. Thus, this is a "reduced-form" version of a more complete model in Appendix A, where I present a demand-based model of variable markups by linking differences in expenditure across the consumer-income distribution to firm-size distribution through product quality. The aim of this simple model is two-fold. First, I derive a general expression of markups under consumer heterogeneity. Second, I derive empirically testable relationship between firms' marginal costs and size, and firms' markups and its size, and present the conditions under which those relationships hold.

Profit maximization. Households are indexed by $h \in \mathcal{H}$ and allocate their income I_h across consumption bundle that comprises of product varieties produced by firm i. Each variety has a quality ζ_i. Each firm produces a unique variety of product, and therefore i indexes both firms and products. Consumer heterogeneity is introduced as following: households vary in their (i) demand elasticities denoted by $\sigma_h > 1$, and (ii) taste for quality $\nu_h > 0$, which follow the assumptions:

Assumption 1. *Household utility from consuming better quality increases with their income levels such that $\nu_h < \nu_{h'}$ if $I_h < I_{h'}$.*

Assumption 2. *Price elasticity of demand is weakly decreasing in income levels, i.e., $\sigma_h \geq \sigma_{h'}$ if $I_h < I_{h'}$.*

Assumption 1 follows from the standard idea in the trade and IO literature that richer consumers are more willing to pay for product quality (Linder 1961). Assumption 2 is verified from the data in Section 3.

Firm Production. Consider a firm i that uses a flexible input factor X to produce output Q_i. The firm's profit function is given by:

$$\Pi_i = P_i Q_i - C(W_i^X(\zeta_i), \Omega_i, Q_i) = \left[P_i - MC(W_i^X(\zeta_i), \Omega_i)\right] Q_i$$

where $Q_i = \Omega_i F_i(X_i)$, P_i is output price, X_i is the input demand, Ω_i is firms' exogenous physical productivity, and $F_i(.)$ is firm's production function. $C(.)$ and $MC(.)$ represent the firm's cost and marginal cost functions, respectively. For simplicity of exposition here, I assume that the production function exhibits constant returns to scale. However, in estimation of markups and marginal costs (described in next section), I allow for flexible returns to scale. Let the price of the input $W_i^X(\zeta_i)$ be a function of output quality ζ_i. I adopt the following two assumptions for differentiated sector:

Assumption 3. *Output quality ζ_i is increasing in firm productivity Ω_i, i.e., $\frac{\partial \zeta_i}{\partial \Omega_i} > 0$.*

Assumption 4. *The price of material inputs is increasing in output quality, i.e., $W_i^{X'}(.) > 0$.*

The assumptions follow the relationship from Kugler and Verhoogen (2011) in reduced-form. First, based on Kugler and Verhoogen (2011) framework, producing better quality output requires better quality input and there are complementarities between firm's productivity and input quality. This is Assumption 3. Second, higher input quality is reflected in higher prices paid by firms to source those inputs. Thus, input prices are increasing in output quality (Assumption 4). Section 3 provides empirical evidence consistent with these assumptions.

Market Structure. We define the market structure as following. Let Q_d is the total quantity of the good in the district d, $\mathcal{I}_d$ is the number of firms supplying that good in the district. Thus, $Q_d = \sum_{i=1}^{\mathcal{I}_d} Q_i$. I define $\varphi_i \equiv \frac{\partial Q_d}{\partial Q_i}$ as the parameter that dictates how much a firm's output affects the total output in the market. Next, I follow the 'conduct parameter' approach to model strategic interactions among firms from Atkin and Donaldson (2015) and define $\Phi_i \equiv \frac{Q_d}{Q_i} . \frac{1}{\varphi_i}$. This allows me to use a single parameter to proxy for competition, instead of separately identifying the underlying parameter. The competitiveness parameter can be summarized under different models of competition: $\Phi_i \to \infty$ under perfect competition; $\Phi_i = \mathcal{I}_d$ under monopolistic behavior; and $\Phi_i = 1$ in collusive environment.

Firm's optimization. A profit-maximizing firm will choose quantity to equate its marginal revenue to marginal costs $MC_i \equiv MC(W_i^X(\zeta_i), \Omega_i)$:

$$\underbrace{\frac{\partial (P_i Q_i)}{\partial Q_i}}_{\text{Marginal Revenue}} = \frac{\partial P(Q_d)}{\partial Q_i} Q_i + P_i = MC_i$$

Dividing both sides of the equation by P_i yields the following expression of markups (P_i/MC_i):

$$\mu_i = \left(1 + \frac{\partial \log P_i}{\partial \log Q_i}\right)^{-1}$$

Under the market structure described above, above markups expression can be rewritten as:[8]

$$\mu_i \equiv \mu(\Phi_1, \tilde{\sigma}_i) = \left(1 + \frac{1}{\Phi_i} \frac{\partial \log P_i}{\partial \log Q_d}\right)^{-1} \tag{1}$$

where the second term in the parenthesis is the inverse of the demand elasticity for the firm, denoted by $\tilde{\sigma}_i$. Let ψ_{hi} be the share of firm's sales made to the consumer group h $\left(\psi_{ih} = \frac{Q_{ih}}{\sum_{h\in\mathcal{H}} Q_{ih}}\right)$. Market clearing on quantities $\left(Q_i = \sum_{h\in\mathcal{H}} Q_{hi}\right)$ links the firm's aggregate demand elasticity to the demand elasticities of its consumer base through the share of sales the firm makes to that consumer base:

$$\tilde{\sigma}_i = \sum_{h\in\mathcal{H}} \psi_{hi} \,.\, \sigma_h \tag{2}$$

Under the set of consumer preferences and firm production, I can establish the sets of results described below. The proof for these propositions is provided in Appendix B.1. Before moving forward, I define assortative matching as the following:

Definition 1 (**Assortative matching**). *The share of firms' sales made to households with higher quality valuation ν_h increases in product quality ζ.*

Proposition 1. *If Assumptions 3 and 4 hold, marginal costs for a firm are decreasing in its size in the homogeneous sector, while they are increasing in size in the differentiated sector.*

Proposition 1 follows from the insight that producing better quality raises the marginal cost due to higher input costs. Because in equilibrium better quality is produced by more productive and larger firms, there is a positive correlation between marginal costs and firm size. In homogeneous goods sector with no quality differentiation, marginal costs are negatively associated with productivity.

Proposition 2 (**Demand-based markup channel**). *Markups are (weakly) increasing in firm size in homogeneous goods sector. Additionally, if Assumptions 1 and 2 hold, then the positive relation of firm size and markup is stronger in the differentiated sector under assortative matching.*

The intuition behind Proposition 2 is that any differences in ψ_{hi} in equation 2 will lead to firms weighing demand elasticities of their consumer base differently. Assortative matching on product quality is one source of systematic variation in ψ_{hi} and makes larger firms in differentiated sector assign larger weights on price elasticities of wealthier consumer base and charge higher markups.

[8] This follows from the relationship: $\frac{\partial \log P_i}{\partial \log Q_i} = \left(\frac{\partial Q_d}{\partial Q_i}\frac{Q_i}{Q_d}\right) \frac{\partial \log P_i}{\partial \log Q_d}$. The term in parenthesis on the right-hand side of the equation is the inverse of the conduct parameter Φ_i as defined above.

2.2 Estimating markups and marginal costs from production data

Equation 1 showed that markups depends on the market supply-conduct and underlying consumer demand, both of which are unobserved. I next describe the "production approach" to retrieve markups from firm production data. The distinct advantage of this approach is its generality. The production approach allows me to measure firms' markups without knowledge of or having to take a stand on many aspects of the theory such as imposing parametric assumptions on consumer demand, or the underlying nature of competition, or assumptions on the returns to scale.

Cost minimization. I follow the cost-minimization approach popularized by Jan De Loecker in his various contributions, technical details for which are provided in Appendix B.2 along with the required set of assumptions. The key insight is that under a static and flexible input, and for which the firm is a input price taker, the markup is identified as the wedge between output elasticities and revenue share of that input. I follow the IO literature and use material inputs as the flexible input in production to compute the output elasticity.[9] The approach provides the following expression for markups μ_{ij} for firm i producing product j:

$$\mu_{ij} \equiv \frac{P_{ij}}{MC_{ij}} = \underbrace{\left(\frac{\partial \log Q_{ij}}{\partial \log X_{ij}}\right)}_{\theta^X_{ij}\text{(Output Elasticity)}} \Big/ \underbrace{\left(\frac{W^X_{ij} \, . \, X_{ij}}{P_{ij} \, . \, Q_{ij}}\right)}_{\alpha^X_{ij}\text{(Expenditure Share)}} \tag{3}$$

where P denotes the price of output Q, W^X denotes the price of flexible input X, MC is the marginal production cost, θ^X is the output elasticity with respect to the flexible input, and α^X is the expenditure on that flexible input as share of firm's revenue. Because more than half of the firms in my data produce more than one product, I specifically follow De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) to estimate markups at firm-product level. Once the markups are estimated for all firm-products, I can obtain marginal costs using information on firm-product prices from the data and dividing it by corresponding markups. Finally, to avoid any potential effect of outliers for our results, I trim the sample at 5^{th} and 95^{th} percentiles of the markup distribution. Appendix Table B.2 shows the average and median markup by sector.

This estimation procedure allows me to overcome two biases in markup estimates relative

[9]In principle, one could use labor as the flexible input similar to previous work on markup estimation. However, there is strong evidence that Indian labor markets are highly regulated and that firms exert monopsony power in the labor market (Brooks, Kaboski, Li, and Qian 2021). This imposes an strong assumption on firms' adjustment costs and input market wedges when using labor as flexible input (see Appendix B.2.3 for a discussion on these issues). As an alternative, I have used electricity as the flexible input in production. The estimates obtained from electricity and material inputs as flexible input measures are strongly correlated (Appendix Figure B.3). For the main analysis, I rely on the estimates obtained using material inputs because data on electricity expenses is missing for about 20 percent of the observations.

to existing work. First, due to data limitations, procedures that uses revenue-based measure of productivity estimation typically rely on industry-level price deflators. This leads to measurement error when firms produce differentiated products, can price differentiate or have market power. This is the 'output price bias' as described in (De Loecker, Goldberg, Khandelwal, and Pavcnik 2016). I use physical output instead of revenue which solves the output price bias. Second, and more specific to my setting, unobserved differences in input quality across firms and over time could generate bias in productivity estimation by inducing an 'input price bias' (De Loecker, Goldberg, Khandelwal, and Pavcnik 2016; de Roux, Eslava, Franco, and Verhoogen 2020). I address the input price bias by adding as controls prices for input factors (wages and materials) and output to the production function. This controls for the unobserved variation in input quality by using information on output prices, with the intuition that input and output prices contain information on both output and input quality (Kugler and Verhoogen 2011).[10]

2.3 Markups and misallocation

The cost-minimization condition with respect to X above also yields the following expression for firms' marginal revenue product of input (MRPX):

$$\text{MRPX}_i \equiv P_i \frac{\partial Q_i(.)}{\partial X_i} = \mu(\Phi_i, \tilde{\sigma}_i)\, W_i^X \tag{4}$$

The marginal revenue product for the input X is directly proportional to markups μ_i (output wedge). Dispersion in markups generates variation in MRPX and is considered a distortion that generates misallocation.

However, as equation 4 depicts, the extent to which markup variation causes misallocation and generates allocative inefficiency depends on sources behind markups. When markups are exogenous, they are isomorphic to input distortions considered in Hsieh and Klenow (2009) and providing subsidies to high marginal product firms would counteract the effect of these wedges by reallocating production factors across firms and increasing aggregate productivity. If, instead, markup dispersion is driven by firms facing heterogeneous price elasticities of demand, then the underlying distortions are endogenous and generated by decisions of optimizing agents to begin with. The price sensitivity of consumers, which determines firms' elasticity of demand through $\tilde{\sigma}_i$, is guided by consumer preferences and is less addressable by policy because it is not susceptible to reallocation. Gains from reallocation would therefore be smaller, and their magnitude will depend on the extent of markup dispersion caused by demand factors.

[10]See De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for a formal model and detailed discussion on the input price bias. Their methodology is also summarized in Appendix B.2.

2.4 Data

1. Firm-level data. The primary data used in this analysis is Indian plant panel-data, the Annual Survey of Industries (ASI) maintained by the Ministry of Statistics. The basic unit of observation in the ASI is an establishment. I use the data from 1998 to 2009 that contain both consistent product level information and establishment location information during these years.[11] The sample frame for the survey is all manufacturing establishments in India that employ more than 10 workers. Establishments with more than 100 workers ("census" establishments) are surveyed every year, while smaller establishments are randomly sampled each year. The data contains establishment-level identifiers across years for both census and non-census establishments, allowing me to construct panel data for both types of establishments.[12] I match the establishment-level panel data to a separate ASI cross-sectional data previously maintained by the Ministry, allowing me to obtain the district in which the establishment is located.[13] The ASI allows owners who have more than one establishment in the same state and industry to provide a joint return, but less than 5 percent of my sample do so, and the analysis is conducted at the level of the establishment. I treat each establishment as a separate firm but the results of the paper hold when I explicitly allow for only single-establishment firms.[14] I limit my analysis to domestic firms by excluding the firms that report non-zero share of their sales exported.

A key advantage of the ASI data is that it provides information on factory-gate wholesale prices for the reporting firms. Indian firms are required by the 1956 Companies Act to disclose product-level information on capacities, production, and sales in their annual reports. This enables tracking firm's product mix over time. Product-level information is available for 80 percent manufacturing firms, which collectively account for more than 90 percent of labor force for the ASI firms.

Firms report products in the ASI survey using ASI Commodity Classification (ASICC) codes which is the most refined level of product available in the data.[15] Table E.1 reports the basic summary statistics by two-digit NIC (industrial classification system for India) sector. Firms in ASI not only report total sales, but also report sales and quantity sold broken down by product. As the product definition is available at highly disaggregated level, unit values are interpreted as prices. I use this information to define per-unit price as (Total Sales Value)/(Total Quantity Sold).

2. Other data. Consumption data for households are from Indian National Sample Survey (NSS)

[11]The ASI uses accounting year which runs from April 1 to March 31. I refer to each accounting year based on the start of the period; for example, the year I call "2000" runs from April 1, 2000 to March 31, 2001.

[12]See Martin, Nataraj, and Harrison (2017) for more details on the ASI data.

[13]A district is an administrative unit in India, with an average of 17 districts per state. A district is comparable to US county in size. On average, a district has approximately 2 million total residents.

[14]Therefore, going forward, I use the term firms which will refer to the establishment.

[15]A product group is the most refined category to which a product belongs in the data. Few examples of product category include cotton shirts, wooden chair, black tea, sugar, cotton yarn. While unit of measurement could vary across groups, all products within the same group are measured in the same units.

conducted between years 1998 and 2009. The survey records total household expenditure and quantity bought by households across 256 product categories, which I use to construct per-unit prices at the household-level. The survey is a nationally representative repeated cross-sectional sample of about 500,000 households with sampling weights provided at the district-level. Weather data collected by the University of Delaware is used to construct a time series of rainfall received across Indian districts since the year 1960. These data are gridded by longitude and latitude lines. In order to match these to districts, I simply use the closest point on the grid to the center of the district and assign that level of rainfall to the district for each year. The agricultural data on district-level cropping patterns, crop prices and crop yields comes from the Ministry of Agriculture.

3 Stylized facts on markup variation and consumer demand

I document facts consistent with *assortative matching* — that is, the tendency of wealthier consumers to source their consumption from goods produced by larger firms —, and demand-based markup channel. I show that (1) larger firms incur higher marginal costs and charge higher markups for their products; (2) the positive relation between firm size, costs and markups is stronger in quality differentiated sectors; (3) richer households consume higher-priced products; and (4) price elasticity of demand is decreasing in household income levels. Together, these facts verify the assumptions made and are consistent with the propositions presented in Section 2.1.

1. Firm-level facts. Panel (a) of Figure I shows the relation between log marginal costs and log number of employees, the closest proxy in the data for unobserved firm productivity.[16] The relationship controls for district-product-year fixed effects to account for any differences, both observed or unobserved, across regions that might contribute to differences in firm costs. The figure shows that within the same narrow product group and located in the same district, smaller firms incur lower marginal costs than larger firms. Specifically, firms with 10 percent larger labor force have 0.41 percent higher costs (Column 1, Panel (a) of Table I).

The positive relationship between marginal costs and firm-size might seem surprising at a first look. In standard production functions, marginal costs are inversely related to physical efficiency which would imply lower marginal costs for larger firms. However, the underlying assumption in those functional form for costs is of constant input prices across firms. This assumption is not valid when firms produce differentiated goods that will require variation in input quality, and therefore, will be reflected in differences in the input prices (Kugler and Verhoogen 2011). Under

[16] Appendix Table E.2 shows that these results, as well as the results that follow, are robust if I use firms' total sales or fixed assets as alternate proxies for size. Labor force is my preferred proxy as unlike sales or physical productivity (which is estimated through the data), it does not induce a measurement error in the independent variable that could be correlated with estimated markups and marginal costs. The positive relationship between the size of firms' labor force and its productivity is documented in Bartelsman, Haltiwanger, and Scarpetta (2013).

this production function, firm's productivity and the input quality are complementarity to each other and marginal costs will increase in firm-size (Proposition 1).[17] The evidence in Columns 3-6 of Panel (a) in Table I (non-parametrically presented in Figure E.1) is consistent with such production function: larger firms use higher priced inputs, are more capital intensive, pay higher wages per-unit labor, and have higher physical productivity (TFPQ).

In fact, Panel (c) of Figure I shows that the positive correlation of marginal costs and markups with firm size is stronger in sectors with greater scope of quality variation, proxied using Rauch (1999) classification of product differentiation, which is identified as products not traded on an organized exchange or listed in reference manuals. The classification is available for SITC 4-digit products, which I concord to the Indian ASI product classification. Panel (b) of Table I reports these correlations. Column 1 shows that the positive relation between marginal costs and firm size, as well as input prices and firm size, is entirely driven by more differentiated sectors. More importantly, as Columns 3-6 of the table show, the underlying factors of marginal costs driving the different correlations with size across the two sectors are input prices (which reflect input quality) and not the differences in the distribution of physical productivity (TFPQ) across the two sectors.

Next, panel (b) of Figure I documents the central findings of the paper: larger firms also charge higher markups for their products. As before, the values on both axes are after controlling for district-product-year fixed effects. This ensures that I am not comparing markups across regions which might differ along unobserved consumer characteristics or market structure. Panel (d) of Figure I shows that the positive correlation of markups with firm size is stronger in sectors with greater quality differentiation. Table I summarizes these correlations. Firms with 10 percent larger labor force charge 0.56 percent higher markups (Column 2, Panel (a)). Column 2 in the bottom panel of Table I shows that positive relationship between firm size and markups is stronger in more differentiated sectors.[18]

2. Household-level facts. Panel (a) of Figure II documents the relationship between per-unit price for a manufactured good consumed by households and their income. The estimates are after controlling for region-by-product fixed effects, where region is either a town or village and is finer geographical unit than a district, which allows to compare price differences within the same product group (e.g., clothes) for households located in a narrow geographical region. I also include controls of households' primary occupation, size, religion and social group which absorbs any observable

[17] An alternative explanation for the positive relationship between costs and firm-size could be decreasing returns to scale. If that were the case, we should not find the relationship to be different across sectors. This is inconsistent with the empirical findings discussed next. Moreover, I estimate returns to scale and do not find evidence to support decreasing returns to scale (see Appendix Table B.1 which shows that the average sum of estimated factor input shares is close to one).

[18] In Appendix C.1, I conduct multiple tests that provide evidence inconsistent with measurement error as a potential driver of these correlations.

Figure I: Firms' markups, marginal costs and size

All variables are measured in logs. The figure shows the relation between firm's per-unit markups, marginal costs and labor force. The top panels shows the average relation by firm-size, and the bottom panel shows the relation by quality-differentiation using the definition in Rauch (1999). The specification controls for district-by-product-by-year fixed effects. Each dot represents 1% of observations.

differences across households that might affect their consumption choices. The evidence shows that wealthier households consume higher-priced products within a narrow definition of a product group.

Next, I estimate the price elasticity of demand across income groups. I rely on the demand curve expression $\log Q_{hgi} = -\sigma_{hgi} \log P_i$ for good i by household h in income group g. Because NSS does not provide with a panel data on households, I estimate price elasticities at the income group

Table I: Baseline Correlations: Firm-size, markups and costs

	Dependent variable: log of ...					
	Marg.Costs	Markups	Input Price	K/L	Wages	TFPQ
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel (a). Average					
(log) labor	0.041***	0.056***	0.063***	0.098***	0.189***	0.150***
	[0.009]	[0.007]	[0.013]	[0.017]	[0.008]	[0.005]
	Panel (b). By quality differentiation					
(log) labor	-0.023*	0.077***	0.051***	0.073***	0.184***	0.155**
	[0.013]	[0.003]	[0.007]	[0.020]	[0.010]	[0.066]
(log) labor ×	0.117***	0.009**	0.019**	0.046**	0.008	-0.010
1(different. good)	[0.016]	[0.004]	[0.008]	[0.020]	[0.009]	[0.074]
Observations	167,221	167,221	443,022	167,221	167,221	167,221
R-squared	0.870	0.638	0.410	0.656	0.803	0.458
Industry f.e.	✓	✓	✓	✓	✓	✓
District-prod.-year f.e.	✓	✓	✓	✓	✓	✓

All variables are measured in logs. The estimates in Panel (a) are from the specification: $\log y_{ijt} = \alpha_k + \alpha_{djt} + \beta \ \log(\text{labor})_{it} + u_{ijt}$, where y_{ijt} is the variable of interest for product j produced by firm i belonging to industry k located in district d in year t. The estimates in Panel (b) are from the specification: $\log y_{ijt} = \alpha_k + \alpha_{djt} + \beta_1 \ \log(\text{labor})_{it} + \beta_2 \ [\log(labor)_{it} \times 1(\text{different. good})_j] + u_{ijt}$. 1(different. good) is a dummy equal to 1 if a product is classified as differentiated. Standard errors clustered at district level are reported in parentheses.

level. This implies that all households within an income group, $\mathcal{H}_g$, have same price elasticity (that is, $\sigma_{hgi} = \sigma_{gi} \ \forall \ h \in \mathcal{H}_g$). Next, for simplicity and technical limitations, I make an assumption that the *relative* price elasticity across income groups is same over the product space. That is, $\sigma_{gi}/\sigma_{g'i} = \sigma_{gi'}/\sigma_{g'i'} \ \forall \ g, g' \in \mathcal{G}$ and $\forall \ i, i' \in \mathcal{I}$, where $\mathcal{G}$ is the set of all income groups and $\mathcal{I}$ is the set of all products that households consume from. The assumption implies while household income groups could differ in their price elasticities across products, the ratio of this difference is same across all income income groups.

Assumption 5. *The relative price elasticity of demand across income groups is same across all product groups.*

Comparing two product varieties i and k gives the following relation between their expenditure (E_{hi}, E_{hk}), and their prices (P_i, P_k):

$$\log \frac{E_{hgi}}{E_{hgk}} = (1 - \sigma_g) \left[\log \frac{P_i}{P_k} \right]$$

where $E_{hgi} = P_i Q_{hgi}$. Under Assumption 5, σ_g can be estimated from the above expression by only considering goods without any quality differentiation.[19] The above equation can be taken to the data using:

$$\log\left(\frac{E_{ihgrt}}{E_{khgrt}}\right) = \alpha_{gir} + \beta_g \log\left(\frac{P_{irt}}{P_{krt}}\right) + \nu_{ihgrt} \tag{5}$$

where r is the region (i.e. town or village) and t is the year of survey. E_{ihgrt} is the household expense on particular product sold at price P_{irt}. As $\beta_g = 1 - \sigma_g$, the above specification provides an estimate of elasticity by income group σ_g. I benchmark $k = 0$ with the most frequent commodity consumed in a region.

The OLS estimate of σ_g will be potentially biased because unobserved taste shocks in the error term could be correlated with price changes. I address this issue by instrumenting local prices $\Delta \log P_{irt}$ with state-level leave-out mean price changes $\frac{1}{N-1}\sum_{k\neq i} \Delta \log P_{krt}$. The instrument identifies the local average treatment effects where the complier group of the instrument will be local and regional sellers for the products. Panel (b) of Figure II shows the estimates of price elasticity of demand across 10 income groups. The price elasticity gradually decreases with income levels before reaching unity for the richest income group. Table E.3 conducts the above exercise parametrically: the price elasticity of demand of the lowest income group quintile is 1.2 times higher than that of the richest quintile.[20,21]

In sum, the evidence presented in this section strongly rejects constant markups across firms. Markups are increasing in firm-size and this relationship is stronger in sectors that are differentiable in product quality. Results from the household consumption baskets show that wealthier households have lower price elasticity of demand. They also consume higher priced goods than poor households within the same product category which suggests they source a higher share of their consumption from larger firms. This evidence is consistent with a *demand-based* markup channel: producing better quality and selling to wealthier, less demand elastic households lead larger firms to incur higher costs and charge higher markups.

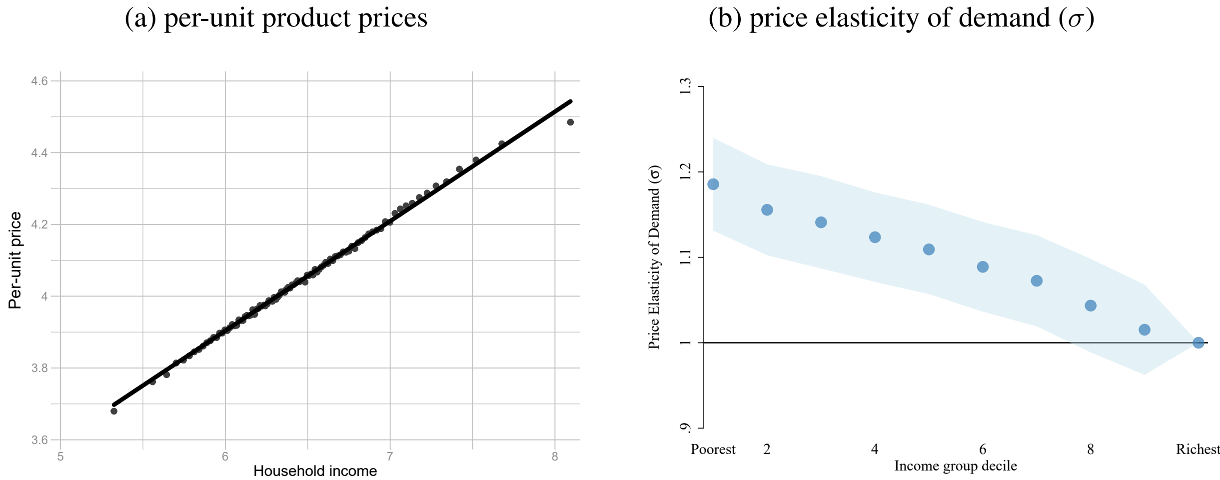
Before proceeding forward, it is important to highlight that it is the combination of being able to obtain *both* marginal costs and markups that provides support for the demand-based markup

[19] This information is sourced from NSS Consumption Survey and includes products recorded under "grains" and "pulses" categories. It includes quantities and prices for rice, wheat, maida, suji, arhar, split gram, whole gram, moong, masur, peas, and soyabean. This is important because quality could be correlated with both demand and prices and thus estimating σ_g based on differentiated products would generate bias in the estimates.

[20] The average estimate of elasticity is comparable to Li (2021) who also estimates the price elasticity in the NSS consumption dataset. In the context of developed countries, Faber and Fally (2020) find low differences of 0.4 in elasticities of substitution across the lowest- and highest-income groups in the US, while Auer, Burstein, Lein, and Vogel (2023) estimate the elasticities for the lowest-income groups to be double that of the highest-income groups.

[21] Table E.4 shows suggestive evidence consistent with Assumption 5: price elasticities are decreasing in household income levels across multiple product groups — including vegetables, fruits, tobacco products, footwear and clothing — that are plausibly more differentiated than grains and pulses.

Figure II: Household income, product prices and price elasticity of demand



Panel (a) shows the relation between log per-unit prices for manufactured goods paid by households and log household income (as proxied by total consumption). The specification controls for product-by-region-by-year fixed effects and household controls (industry of occupation, type of occupation, size religion and social group). Each dot represents 1% of observations. Panel (b) reports the estimates of price-elasticity of demand across income groups (σ_g) based on the estimating equation 5. The estimates are based on a IV-2SLS specification that instruments price of a good with state-level leave out mean price changes. 95% confidence intervals are represented by shaded blue area. Source: NSS.

channel. While the observation that markups are increasing in firm-size is independently made — either through direct or indirect evidence — in recent work (Dhingra and Morrow 2019; Edmond, Midrigan, and Xu 2019), the relation between costs and firm-size is negative. The results above show that while these models fit well the correlations documented in the homogeneous sector, they are inconsistent with the relations in the differentiated sector. Similarly, in supply-side models of variable markups such as Atkeson and Burstein (2008); Edmond, Midrigan, and Xu (2015) where consumers have CES preferences and firms compete in imperfectly competitive environment, markups are higher for larger firms. The results on price elasticity of demand across consumers and positive relation between cost and firm-size both do not fit well these models. Appendix D discusses these and other alternative models of firm-heterogeneity, and Appendix Table D.1 compares the correlation between firm-size, markups and costs as made across these frameworks.

4 Isolating the role of consumer demand for markup variation

The equilibrium relationship between markups and firm size documented in Section 3 does not identify the causal effect of demand composition. Equation 1 (and Proposition 2) suggests that larger firms could charge higher markups in equilibrium because they have larger market shares or because they face variable elasticity of demand. Moreover, variable elasticity of demand could arise from sources other than firms' demand composition.[22] To isolate the role of demand composition for markups, I incorporate demand shocks to the poorest households in the framework presented in Section 2.1 to generate the following testable prediction.

Proposition 3. *Firms lower their markups in response to an increase in demand from poor households. Additionally, the markup response is convex (U-shaped) with respect to the firm size.*

Proof. See Appendix B.1.

To test this prediction, I next propose a research design that uses quasi-exogenous changes to consumer demand across the income distribution. These changes to demand from poor households affect the demand composition of firms differently because consumers across income levels differ in the shares of their consumption basket sourced from large-, mid-, and small-sized firms. I then study how firms change their markups in response to changes in their demand composition.

4.1 Empirical Strategy

The objective is to understand how firms adjust their markups in response to changes their demand. The equilibrium relation between price P_{ijt} and quantity Q_{ijt} for firm i and product j is given by:

$$\log P_{ijt} = \alpha_0 + \alpha_1 \log Q_{ijt} + \nu_{ijt}$$

Using the identity $\log P_{ijt} = \log \mu_{ijt} + \log MC_{ijt}$, the above relation can be rewritten as:

$$\log \mu_{ijt} = \alpha_0 + \alpha_1 \log Q_{ijt} + (\nu_{ijt} - \log MC_{ijt}) \tag{6}$$

Estimating 6 using ordinary least squares (OLS) methods could lead to biased estimates of α_1. Any correlation between markup and quantities will not identify the causal effect of demand on markups because of (i) reverse causality: higher priced products (the ones with higher quality) could observe an increase in their demand, that is causality might run from markups to quantities; (ii) omitted variable bias: changes along the demand curve, i.e. changes to marginal costs of production, could change firms' markups and therefore the demand Q_{it}; and (iii) measurement error: estimates could be mechanically negative as prices are calculated as product revenue divided by its quantity sold. A solution to this is to obtain an exogenous demand shifter to Q_{ijt} that is unlikely to be correlated with the firms' marginal cost and the market structure. I propose one such instrument for changes in firm's demand: changes to consumer income due to local rainfall fluctuations.

Intuition behind Identification. Similar to many other developing countries, majority of the poor in India are employed in the agricultural sector. About 66 percent of males and 82 percent of

[22] For example, see theoretical work by Zhelobodko, Kokovin, Parenti, and Thisse (2012); Edmond, Midrigan, and Xu (2019); Dhingra and Morrow (2019).

females in rural India report agriculture as their principal economic activity (Mahajan and Gupta 2011).[23] More than two-third of farmed area in India is rain-fed; and thus agricultural production and rural income are considerably dependent on rainfall. Rainfall exhibits significant variation across districts and over years, generating income changes for poor households in those districts. More importantly, these weather-induced changes to income are *transitory* in nature and lack any persistence across years even within districts (Table E.7). These income changes over years affect the demand for firms that cater more to the poor households than firms that cater less to them. To see this, notice that quantity Q_{hijt} demanded by a income group h over time t is a function of the prices P_{ijt}, the price index P_{ht} faced by the group, income I_{ht} for that group, and other factors ν_{hijt}:

$$Q_{hijt} \equiv D(P_{ijt}, P_{ht}, I_{ht}, \nu_{hijt}) \tag{7}$$

The derivative with response to the third argument is $D_3 > 0$ which implies an exogenous income shifter for group h will increase the demand Q_{hijt} from that group.[24] The aggregate demand for firm i is the sum of its total sales to each consumer group, i.e. $Q_{ijt} = \sum_h Q_{hijt}$. Assortative matching dictates that firms across the size distribution — more so in the quality differentiated sector — differ in their share of sales made to different income groups, and therefore, differences in income changes across consumer groups affect the demand Q_{ijt} for firms with varied intensities.[25]

Main Specification. To estimate how rain shocks affect firm outcomes, I run the following specification:

$$\log y_{ijt} = \beta \text{ Shock}_{dt} + \alpha_{ij} + \alpha_{jt} + \gamma \tilde{X}_{ijt} + \epsilon_{ijt} \tag{8}$$

where y_{ijt} is the year t outcome of interest (demand, quantity sold, costs, and markups) for product j produced by firm i located in district d. Shock_{dt} are local rain shocks as defined below. As products produced by different firms could differ across various characteristics, I include firm-product fixed effects α_{ij} which absorbs any time-invariant firm-product unobservables (for example, any constant quality differences). The presence of product-year fixed effects α_{jt} controls for product-specific inflation and any macro-economic shock at the product level. $\tilde{X}_{ijt}$ are set of firm and market level controls described as they are used in Section 4.2. The reduced form coefficient β in the

[23]The relationship between agricultural employment and income levels across districts is evident from Figure E.2 which shows that average income in the district is systematically decreasing in its share of population employed in the agricultural sector.

[24]For the ease of exposition, I have abstracted away from presence of household savings. In reality, it is possible that households might smooth their consumption by saving more in response to transitory income shocks. However, as discussed later in the identification assumptions, the evidence on marginal propensity to consume in response to transitory income changes directly refutes this possibility.

[25]Using weather driven income changes has an additional advantage over other measures of local income changes that could be driven by changes in aggregate price levels (for e.g., industry level wage growth). To see this, we can decompose $\Delta \log I_{ht}$ into a function of aggregate prices $f(P_{ht})$ and a residual variation independent of prices ϵ^I_{ht}: $\Delta \log I_{ht} = f(P_{ht}) + \epsilon^I_{ht}$. Rain shocks have the advantage of affecting the residual variation ϵ^I_{ht}.

specification is straightforward to interpret as the elasticity of the response of firm-product level outcomes to rain shocks.

Following the non-linear relationship between local rainfall deviations in a year and agricultural yields in Figure E.3, I define a positive shock if the annual rainfall measure is above the 70^{th} percentile and negative shock as rainfall measure below the 30^{th} percentile within the district. The positive and negative shocks should not be taken in an absolute sense as I am not comparing districts that usually receive higher rainfall to those that usually receive lower rainfall. This measure simply captures high or low-rainfall years for each district during 1960-2009. For the analysis, I follow Jayachandran (2006) and define rain shock as equal to +1 for positive shock, -1 for negative shock, and 0 otherwise. The mean value of the rain shock measure is -0.14 with standard deviation of 0.78. Columns 1-2 of Table E.5 show the effect of rain shocks on local agricultural outcomes: positive rain shocks increase crop yields in the district by 5 percent and revenue by 3.5 percent.[26]

Identification Assumptions. Consistent estimation of β in specification 8 requires two conditions to be satisfied: relevance of rain shocks, that is, Shock_{dt} and $\log Q_{ijt}$ should be correlated; and exclusion restriction, that is, Shock_{dt} is uncorrelated with ϵ_{ijt}. Relevance can be directly tested in the data — local rainfall deviation should be strongly correlated with the local income and the quantity demanded for poor households. Two results lend strong support to the hypothesis that rain shocks change the *relative* demand of the poor households, and affect the demand disproportionately across the firm-size distribution.

First, I show that rain shocks affect the wages of population employed in agriculture. Column 3 of Table E.5 shows the effect of rain shocks on incomes of the poor: daily wages in agricultural sector increase by 2.7 percent. Rain shocks do not affect wages for households employed outside agricultural sector or for non-rural labor force (Columns 4 and 5). Next, I document that poor households have higher marginal propensity to consume (MPC) out of temporary income changes. Figure E.5 reports the distribution of MPC across the income distribution.[27] For same increases in income (and conditional on prices), quantity demanded increases more for the poor population. Taken together, these results provide strong support that rain shocks generate significant variations in demand for the poor households.

Second, I check how rain shocks affect firms' idiosyncratic demand. Following the influential

[26] I run the following specification: $y_{dct} = \beta \times \text{Shock}_{dt} + \alpha_{dc} + \alpha_{ct} + \epsilon_{dct}$, where the outcome variable is either average yield (output per hectare) or revenue for crop c across fifteen major crops in India and α_{dc} and α_{ct} are the district-crop and crop-year fixed effects.

[27] I follow Gruber (1997) and calculate the MPC using the observed drop in consumption upon unemployment. Using a monthly panel data on 100,000 households from CMIE household consumption data, I estimate the following regression for household h in town v at month t: $\Delta \log E_{hgvt} = \alpha_g \, \Delta \log I_{hgvt} + \beta_h + \beta_{vt} + \epsilon_{hvt}$, where β_h is the household fixed effect, β_{vt} is a town-year fixed effect that captures the total resources available in the town-month and aggregate shocks in month t, and g is the income group. As the regression is run on a panel data at household-month level, the coefficient α_g is identified of the variation in *within* household income across months.

work by Foster, Haltiwanger, and Syverson (2008), I use the production data and obtain firm's idiosyncratic demand by isolating total quantity movement from quantity movement due to a change in supply-side change in prices. Specifically, I estimate firm-product level demand-shifters η_{ijt} using:

$$\log Q_{ijt} = \gamma \log P_{ijt} + \alpha_{jt} + \eta_{ijt} \tag{9}$$

α_{jt} absorbs yearly changes at the product-level (in both supply and demand), and η_{ijt} are firm-product demand shifters. Estimating equation 9 could lead to positive bias in estimates of price elasticity γ, because firms could respond to demand shifters η_{ijt} by increasing prices. I overcome this problem by using changes to marginal cost as instrumental variables (IV) for supply-side price shifters. Marginal costs incorporate firms' idiosyncratic cost-shifters through changes in their input prices and firm's technology.[28] Thus, it has explanatory power over firms' prices which are unlikely correlated with short-run changes in demand. The demand estimates γ from specification 9 are shown in Table E.6. The IV estimates of elasticities are negative, range from -4.5 to -1.9 across the industries, and are 2 to 4 times more elastic than OLS estimates, consistent with the upward bias due to simultaneity in the OLS estimates.

I use the residuals η_{ijt} from the demand function estimation to provide evidence on the relevance of rain shocks for firm-level demand. Table II reports the correlation of demand shocks η_{ijt} with rainfall shocks (using specification 8). Column 1 shows that firms' estimated idiosyncratic demand increases by 1.2 percent during years of positive rain shocks. Similar result is obtained if I instead use quantity sold by firms as a direct measure of firms' demand (Column 2). Next, I estimate the effects of rain shocks on firms' demand across quartiles of firm-size distribution using:

$$\log y_{ijt} = \sum_{r=1}^{4} \beta^r . \left(\text{Shock}_{dt} \times T_i^r\right) + \alpha_{ij} + \alpha_{jt} + \gamma \tilde{X}_{ijt} + \epsilon_{ijt} \tag{10}$$

where $r \in \{1, 4\}$ indexes each of the four quartiles of the size distribution and T_i^r are dummy variables taking the value of 1 when firm i belongs to quartile r.[29] Panel (a) and (b) of Figure III shows the effects of rain shocks on firms' demand across the size distribution. The effects are strongest for smallest firms and for firms in the middle of the size distribution, and gradually decrease to zero for the largest firms.

Exclusion Restriction. The second identification condition that rain shocks should satisfy is

[28] Relative to Foster, Haltiwanger, and Syverson (2008), I use marginal costs instead of physical productivity (TFPQ) as an instrument for prices. Unlike TFPQ which is estimated at firm-level, marginal costs are estimated at firm-product level and provides greater variation. See Appendix B.1.1 for the discussion of log-separability of marginal costs into TFPQ and input prices.

[29] I define these quartile using firm size (using firm's first occurrence in the panel) based on its labor force relative to two-digit industry average. Using 2-digit industries instead of products increases the number of observations within each quartile and reduce the noise associated with misclassfication.

Table II: Effect of rain shocks on firms' idiosyncratic demand and marginal costs

	Dependent variable:					
	Demand Shifter	log quantity	log of ... marg. cost	TFPQ	wage	input price
	(1)	(2)	(3)	(4)	(5)	(6)
Shock_{dt} (-1/0/+1)	0.012**	0.014***	-0.004	-0.011	0.001	0.001
	[0.005]	[0.005]	[0.007]	[0.008]	[0.002]	[0.008]
Observations	133,094	133,094	133,094	59,965	102,541	239,100
R-squared	0.898	0.975	0.952	0.887	0.922	0.931
Firm f.e.				✓	✓	
Firm-product f.e.	✓	✓	✓			✓
Product-year f.e.	✓	✓	✓	✓	✓	✓

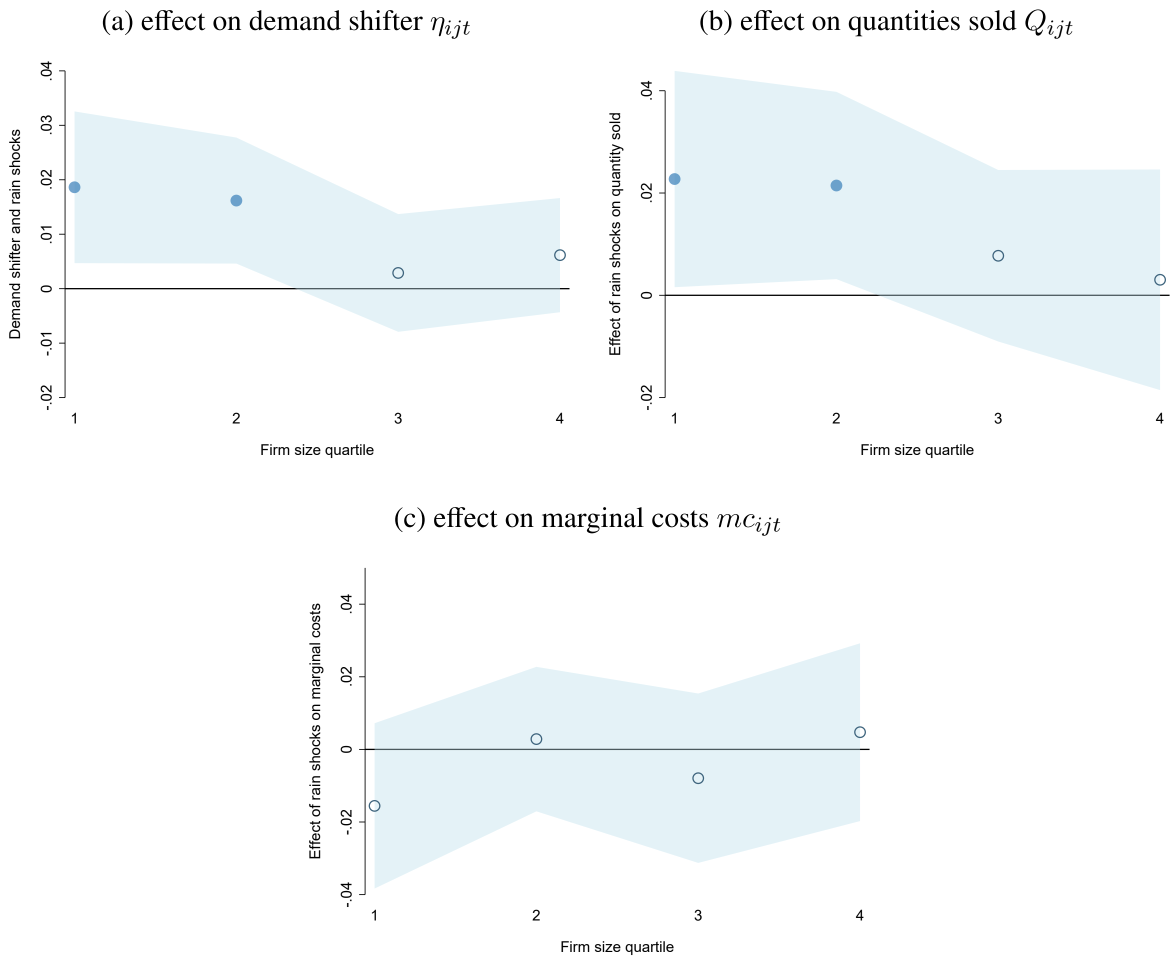
All dependent variables but Column 1 are in logs. Shock_{dt} is defined as +1(-1) if the rainfall in the monsoon months is above(below) the 70^{th}(30^{th}) percentile of the district's usual distribution for monsoon rainfall. It takes the value of 0 if the rainfall is between 30^{th}-70^{th} percentile of district's usual distribution. Standard errors are clustered by district level are reported in parentheses.

exclusion restriction. That is, rain shocks should affect markups only through changes to demand curve faced by firms, and not due to changes to the supply curve.[30] While this assumption cannot be directly tested, I believe the richness of the production data allows me to test whether rain shocks might affect firms' supply curve. As mentioned before, observing prices at firm-product level provides estimates of marginal costs along with markups, allowing to test whether (and how) rain shocks affect marginal costs across firms. Columns 3-6 of Table II report the correlations of rain shocks with firms' marginal costs and its underlying components. I do not find any evidence that rainfall shocks affect marginal costs on average, or firms' physical productivity (TFPQ), wages and prices of material inputs. Figure III Panel (c) shows that rain shocks do not affect marginal costs across the firm-size distribution, and Figure E.4 shows that there is no affect of rain shocks on TFPQ, wages, input prices, and fixed capital across small, medium and large firms.[31]

[30]For example, Asher and Novosad (2012) and Colmer (2021) argue that local rain shocks could affect the manufacturing firms through capital and labor reallocation, respectively, from or to the agricultural sector. However, using a dataset similar to this paper, Santangelo (2019) finds that rain shocks mostly affect local manufacturing firms through a local demand channel, and find little evidence for supply side factors affecting firms' hiring decisions.

[31]In addition to the above identification assumptions, a separate assumption that needs to be satisfied is that an increase in income for poor households could should not decrease their long-run price elasticity. Note that is not in violation of exclusion restriction but could still independently affect markups. For example, poor households can become less price-elastic if higher income in the current year due to better rainfall is predictive of higher income in the future years. As discussed later in section 4.2, a decrease in demand elasticity in years of positive rainfall shock should led to a *increase* in markups. However, I find that markups *decrease* in years of positive rain shocks and thus this mechanism should bias, if anything, the estimates towards zero. I also test for serial correlation of rainfall within districts because serially correlated rainfall shocks could induce permanent shifts in the price-elasticity of demand. Table E.7 shows an absence of any serial correlation in rain shocks.

Figure III: Effect of rain shocks on demand and costs (across firm-size distribution)



The figure shows the estimates of the effect of rain shocks on firms' demand and marginal costs across the firm-size distribution based on specification 10. All specifications control for firm age and size quartile-year fixed effects. 95% confidence intervals are represented by shaded blue area. Bold circles indicate results that are significant at the 10% level, and hollow circles statistically insignificant from 0 at the 10% level.

While rain shocks do not affect marginal costs on average, they could still have a non-zero effect on costs for some firms. For example, an increase in demand could affect costs through changes in X-inefficiencies for few firms and not others. These changes in firms' costs could have an independent supply-side effect on markups and generate bias in estimated β. Therefore, I control for marginal costs in specification 8 in order to isolate markup responses due to changes in demand from rain shocks. This addresses any omitted variable bias by absorbing any component in the error term that might be correlated with both markup changes and quantity produced.

4.2 Results

Table III presents the main results on how average markups respond to rain shocks. Column 1 shows that firms lower their markups by 0.5 percent in years of positive rain shocks. In Columns 2-8, I show that the results are robust to inclusion of various controls. Because multi-plant firms might be less responsive to local shocks, I restrict the analysis to only single plant firms in Column 2. I include controls for firms' age in Column 3 because firm size and age are heavily correlated and thus estimates could confound age- and size-effects. In Column 4, I include controls for firms' size quartile and its interaction with year fixed effects to allow for differences in aggregate shocks across size groups. Column 5 includes controls for past two-years of rain shocks in the district to allow for any effects from lagged changes in demand. In Column 6, I control for market access measure constructed from Allen and Atkin (2016), which is a weighted average rainfall deviation for each district d' connected to district d , where the weights are proportional to the distance between the two districts. Column 7 controls separately for an in-state and an out-state market access measure to allow for separate impact based on whether other districts d' are in the same state as district d or outside the state. Finally, in Column 8 I allow for combined effect of controls from Columns 2-7. As can be seen, addition of these controls has no significant effect on the estimate of average effects of rain shocks on markups.[32]

Table III: Average effect of rain shocks on firms' markups

	Dependent variable: log markup							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Shock$_{dt}$ (-1/0/+1)	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005**	-0.004*	-0.004*
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Observations	133,094	122,828	133,094	133,094	133,094	133,094	133,094	133,094
R-squared	0.989	0.990	0.989	0.989	0.989	0.989	0.989	0.989
Firm-product f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Product-year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Controls	Baseline Specification	Single-plant firms	+ Age control	+ Size-year control	Past 2-year shocks controls	National Market access control	In + out-state market access	(3)-(7) controls

The table reports the average effects of rain shocks on markups, based on specification 8. Shock$_{dt}$ is as defined in the text. All columns include firm-product, product-year fixed effects and control for log marginal costs. Standard errors clustered by district level are reported in parentheses.

Mechanism. Next, I provide evidence supporting Proposition 3 on the role of consumer hetero-

[32]The significant relationship of markups and the null effects on marginal costs remain robust to various specifications of rain shocks (Table E.8). I also analyze the effects of rain-induced local demand shocks on exporters. Markups for exporters are a function of the demand that they face in export markets, rather than the local demand. Therefore, exporters should largely be unaffected by the demand changes from rain shocks. However, if rain shocks were indeed common supply shocks to firms, we would expect then to affect firm costs. Table E.9 shows that neither markups or marginal costs are affected by local rain shocks for exporters.

geneity in driving markup variation. To see this, notice that the markup elasticity to firm i's idiosyncratic demand shocks η_{it} is given by:

$$\underbrace{\frac{d\log\mu_{it}}{d\log\eta_{it}}}_{\text{markup change}} = \frac{-1}{\Phi_{it}\tilde{\sigma}_{it}-1} \times \left(\underbrace{\left[\frac{d\log\Phi_{it}}{d\log Q_{it}}\right]}_{\substack{\text{change due to conduct}\\ \text{(competition effect)}}} + \underbrace{\left[\frac{d\log\tilde{\sigma}_{it}}{d\log Q_{it}}\right]}_{\substack{\text{change due to }\Delta\text{ slope}\\ \text{(slope effect)}}} \right) \times \underbrace{\left[\frac{d\log Q_{it}}{d\log\eta_{it}}\right]}_{\substack{\text{change due to }\Delta Q\\ \text{(size effect)}}} \tag{11}$$

Equation 11 shows that the necessary condition for firms to change its markups in response to a change in demand is that a shift in firms' demand (change due to Δ Q, that is, the *size effect*) is accompanied by either a shift in the slope of firms' demand or due to a change in competition faced by the firm.

So how does one separate the sources behind markup changes? According to Proposition 3, if markup responses to demand shifts are due to changes in demand composition, then an increase in demand from the poor households increases the demand elasticity *only* for firms that sell to both rich and poor households, forcing them to lower their markups. Under assortative matching, these firms are proxied in my data by firms in the middle of the size distribution. Therefore, changes to demand composition, and hence markups, should be strongest for firms in the middle of the size distribution. Smallest firms cater largely to poor — and therefore, homogeneous — consumer base, implying that rain shocks should not affect their demand composition and markups $\left(\frac{d\log\tilde{\sigma}_{it}}{d\log\eta_{it}} \to 0\right)$. The consumers for largest firms are rich households, and therefore, rain shocks do not affect the demand for these firms $\left(\frac{d\log Q_{it}}{d\log\eta_{it}} \approx 0\right)$. That is, in the context of equation 11, the slope effect for smallest firms and the size effect for largest firms are zero which implies that markups should not change for these firms. More importantly, this non-monotonic pattern should only be present in quality-differentiated sectors.

If instead the reduction in markups from higher demand is due to an increase in competition among firms, then the effects should be strongest for the smallest firms which observe the largest increase in demand (but no significant changes to their demand composition). Table D.1 provides a comparison of how markups will change across the firm-size distribution under different models of variable markups proposed in the literature. The table shows that it is only under the demand composition channel, that markup responses to demand shocks will be non-monotonic across the firm-size distribution.

To test this prediction, I estimate the effect of rain shock on markups across quartiles of firm-size distribution using specification 10, estimation results for which are plotted in Figure IV. As the figure shows, rain shocks only affect markups in the middle of the size distribution. The estimates are reported in Table E.10. The coefficient of -0.7 to -0.9 percent and -0.5 to -0.8 percent in the second and third quartile, respectively, of the size distribution is more than two to three times larger than the lowest quartile (which are insignificant across all specifications). Firms in the largest size quartile also do not change their markups. The estimates remain stable after inclusion of various controls from Table III (Columns 2-8).[33] Next, I test whether firms producing more differentiated goods change their markups more by estimating the following specification:

$$\log \mu_{ijt} = \sum_{p\in\{0,1\}} \sum_{r=1}^{4} \beta_p^r \cdot (\text{Shock}_{dt} \times T_i^r \times (Z_{ij} = p)) + \alpha_{ij} + \alpha_{jt} + \Gamma' X_{ijt} + \epsilon_{ijt} \tag{12}$$

where Z_{ij} takes the value of 1 for firms in sectors with greater scope for quality differentiation. If differences in taste over quality are driving the assortative matching, we should observe that the non-monotonic pattern of markup responses should be more prominent in more differentiated sectors. Panel (a) of Figure E.7 shows that the results are consistent with this interpretation.

An issue with the above interpretation of the null responses among large firms is that the customer base for these firms could be spread across the state or the country, instead of being local. If this was indeed the case, then markup responses should be similar across all sectors irrespective of whether the goods are traded locally or nationally. While there is evidence suggestive of high trade costs leading to localized markets for Indian manufacturing firms (e.g., Leemput 2016; Rotemberg 2019), I still analyze the above hypothesis by estimating markup responses by intensity of tradability associated with a product. I follow Mian and Sufi (2014) and create industry-level classification of tradability that relies on the observation that in equilibrium the production of non-tradable goods tend to be spatially more distributed across the country. I create a concentration index using sector's share of national employment. As an example, this definition classifies concrete manufacturing as non-tradable whereas manufacturing of car parts is classified as tradable.[34] Panel (b) of Figure E.7 shows that the results are stronger in industries classified as non-tradable than those classified as tradable.

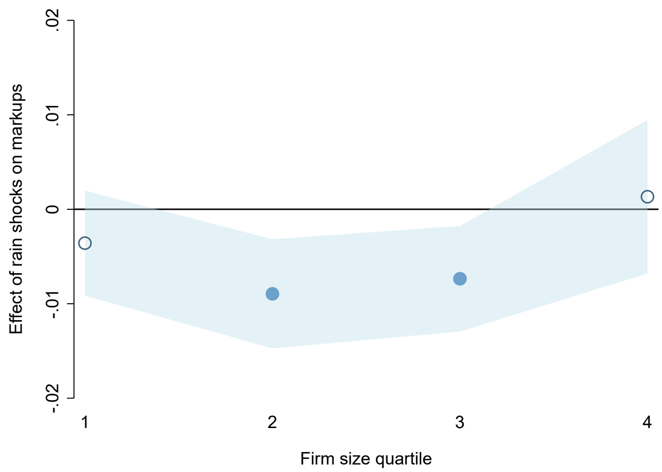
4.3 Alternative mechanisms

In Appendix C.2, I consider alternative explanations for lower markups in periods of increased demand. First, under imperfect competition, incumbent firms could decrease markups due to entry of new firms, or introduce new products in response to higher demand. Second, firms might collude in setting markups and the incentives to deviate from such collusive agreements could increase

[33] In Panel (a) of Figure E.6, I conduct falsification tests using rain shocks realized in the next year rather than the current year. Markups are not responsive to these placebo shocks. I also do not find any evidence for past demand responses (observed from last year's rain shocks) having any persistent effects on future markups (Panel (b) of E.6).

[34] This definition is constructed using the 2005 Economic Census which surveys every national non-farm establishment and records the industry and number of employees. To reduce any noise in the middle of the distribution, I only report results for top and bottom tercile of the tradability classification.

Figure IV: Effect of rain shocks on markups across firm-size distribution



The figure shows the estimates of effects of rain shocks on markups across the firm-size distribution. The specification includes firm-product, product-year fixed effects and controls for firm age, size quartile-year fixed effects, and log marginal costs. 95% confidence intervals are represented by shaded area. Bold circles indicate estimates significant at the 10% level, and hollow circles statistically insignificant from 0 at the 10% level.

when demand increases. Third, consumers might increase their shopping search intensity when their income increases. Fourth, financially constrained firms could raise markups when facing negative demand shocks. A common distinction between these explanations and the demand-based markup channel is that the observed non-monotonic pattern of markup responses across the firm-size distribution documented above is unique only to the latter. I nevertheless examine each of these explanations separately and find empirical evidence inconsistent with any of them.

Altogether, the evidence above shows that differences in demand composition across firms — arising from assortative matching on quality between firms and consumers — are necessary to rationalize the patterns of markup dispersion observed in the data. However, these results leave two related questions open. First, they do not indicate how large are the misallocation losses due to variable markups. Second, from these results, no conclusion can be drawn on the quantitative contribution of demand- and supply-side factors for misallocation losses. In the next section, I address both questions by providing an approach to estimate gains from reallocation under variable markups, and quantify the losses arising separately due to demand- and supply-side factors.

5 Aggregate Implications

In this section, I assess the implications of demand-based markup dispersion for aggregate productivity. Dispersion in markups could arise through differences in nature of competition faced by firms (supply-side factors), or due to differences in consumer preferences (demand-side factors), or a combination of both. Consumer preferences, however, are not susceptible to reallocation and the aggregate productivity gains that could be attained from a hypothetical reallocation exercise will, of course, be correspondingly smaller. In the end, any exercise computing reallocation gains is specific to the underlying model or the hypothetical exercise. For example, Hsieh and Klenow (2009) propose dispersion in revenue productivity and marginal products on input as a measure of allocative inefficiency. The gains from reallocation in their framework is proportion to the variance of the dispersion in revenue productivity (TFPR).

I take a different, yet complementary, approach and ask "how much aggregate productivity gains can be achieved if we remove underlying distortions through a tax-subsidy policy?".[35] Such an approach would need two objects to estimate gains from reallocation. First, it requires an expression that relates aggregate productivity gains to the underlying distortions. For this I rely on a first-order approximation for aggregate productivity growth from Petrin and Levinsohn (2012). An advantage of this approach is that it is general and does not impose structure on underlying demand or market structure. This is especially relevant given that any model-specific exercise for computing reallocation gains will only be as credible as the underlying model of demand and supply, making such an exercise susceptible to model mis-specification (Haltiwanger, Kulick, and Syverson 2018; Asker, Collard-Wexler, and De Loecker 2014) or measurement error (Bils, Klenow, and Ruane 2018; Gollin and Udry 2021).

Second, it requires a tax-subsidy policy to counteract the underlying distortions. I consider a policy that serves a planner's objective to equalize marginal revenue products (MRP) across firms within industry under a fixed aggregate supply of resources, while (erroneously) assuming that any variation in MRP across firms arises only due to presence of exogenous distortions. These conditions are standard in the static misallocation literature. The main result from the exercise is that when markup distortions are endogenous, firms could adjust their markups in response to tax-subsidy policies. This substantially lowers the gains from the intended reallocation policy. For this exercise, I make the following assumption:

Assumption 6. *The slope of demand faced by firms in homogeneous goods sector is constant, but can vary across firms in differentiated sector due to differences in firms' demand composition.*

Assumption 6 follows from the results in Section 3 and 4: differences in demand composition arising from assortative matching between firms and consumers on quality generates provides firm with additional market power in differentiated sectors. The assumption implies that firms in

[35] I consider the policy of optimal subsidy for two reasons. First, it provides a set of simple, easy-to-act rules that can be targeted by policies, based solely on divergence between market prices and social costs. As Dixit (1985) notes: "a distortion is best countered by a tax instrument that acts directly on the relevant margin. Once the relevant margin has been traced, a tax-subsidy policy can be imposed to close the gap." Second, it is also consistent with the development and growth literature which considers misallocation to exist if a planner could implement budget-neutral taxes and subsidies to induce the reallocation of inputs across firms.

homogeneous sector face same slope of demand curve. That is, $\sigma_i = \sigma^{\text{non-diff}}$ in homogeneous sector, and differences in markups across firms in homogeneous sector are thus driven *only* by differences in competitiveness Φ_i. For firms in differentiated sector, however, variation in markups is driven by differences in slope of demand as well as competitive index. Under this assumption, differences in markups in homogeneous and differentiated goods sector can be used to obtain estimates of σ_i and Φ_i for all the firms. This allows to estimate reallocation gains under CES versus variable demand, and under different underlying market structure faced by firms.

5.1 Analytical framework

Aggregate productivity growth. In this subsection, I describe the aggregate productivity growth decomposition from Petrin and Levinsohn (2012). The change in aggregate productivity for a sector s is the difference between changes in output and input costs within that sector:

$$dAP_s = \sum_{i \in I_s} P_i dQ_i - \sum_{i \in I_s} W_i^X dX_i$$

where Q_i is the gross output of firm i, and P_i is firm's price, I_s is the number of firms in sector s. I define the total productivity growth in the economy as the weighted average of sector-level productivity growth: $dAP = \sum_s \gamma_s \ . \ dAP_s$, where γ_s is the share of total output in the economy coming from sector s.[36] As the expression is similar across all sectors, for convenience I omit the notation s going forward. Setting aside firms' entry and exit, the aggregate productivity growth can be decomposed into a within-firm productivity improvement ("technical efficiency") term and an across-firm allocation ("reallocation") term.[37]

$$\text{APG} = \underbrace{\sum_i \lambda_i d\log \Omega_i}_{\text{APG (within)}} + \underbrace{\sum_i \lambda_i \left(\theta_i^X - \alpha_i^X\right) d\log X_i}_{\text{APG (reallocation)}} \tag{13}$$

where θ_i^X is the output elasticity with respect to the input, α_i^X is the input expenditure as share of firm's revenue, Ω_i is firm's technical efficiency, and $\lambda_i \equiv \frac{P_i Q_i}{\sum_i P_i Q_i}$ is firm's (Domar 1961) weight. The output elasticity θ_i^X is obtained by estimating the production function as described in Section 2. The revenue share of firms' input expenses α_i^X and Domar weights λ_i are obtained directly from the data.

[36] Following the literature, I use 4-digit NIC industry classification to define the sectors. There are 125 sectors in the data.

[37] As described in Section 2.4, while the ASI data used in this paper provides consistent firm identifiers across years, it only surveys about one-third of firms in consecutive years, making it difficult to identify the contribution of firms' entry and exit to aggregate productivity growth.

Reallocation policy. Equation 13 shows that reallocation gains are directly related to $d\log X_i$. From Equation 4, the firms input demand X_i is a function of markups μ_i and exogenous distortions τ_i^X (that is, $X_i \equiv X(\mu_i, \tau_i^X)$). The implies that the change in input demand to a tax/subsidy S_i is:

$$d\log X_i \;\; = \;\; \left[\frac{\partial \log X_i}{\partial \log \mu_i}\frac{\partial \log \mu_i}{\partial \log S_i} + \frac{\partial \log X_i}{\partial \log \tau_i}\frac{\partial \log \tau_i}{\partial \log S_i}\right] d\log S_i \tag{14}$$

I define firms' pass-through rate Γ_i as the elasticity of firm's price to its costs as $\Gamma_i \equiv \frac{\partial \log P_i}{\partial \log S_i} = \left[1 + \frac{\partial \log \mu_i}{\partial \log S_i}\right]$.[38] Substituting for this expression in 14, and with some algebra, yields the following relationship between input demand and subsidy (see Appendix B.3 for a detailed derivation):

$$d\log X_i = -\left[\frac{\Gamma_i}{\theta_i^X - \alpha_i^X}\right] d\log S_i \tag{15}$$

There are two factors that affect resource reallocation across firms in response to a subsidy. First, firms with low-demand elasticities pass-through only a fraction of those subsidies into their prices. This fraction is dictated by the pass-through rates for firms. Second, conditional on limited pass-through firms facing low demand elasticities witness less changes in their quantities demanded, and therefore, change their input demand by less.

Equation 15 shows that the reallocation gains can be estimated for a proposed tax-subsidy policy S_i. I obtain such a policy by considering a social planner with the following objective and constraint: (i) the planner equalizes marginal revenue products for inputs (or, alternatively, markups) across firms within a sector; (ii) the planner faces a fixed supply of aggregate factors. These conditions are standard in the static misallocation literature. When dispersion in marginal products is assumed to be exogenous, the expression of tax/subsidy for firm i takes the following form (see Appendix B.3 for details):

$$d\log S_i = \left[\left(\sum_i \left(\frac{\tilde{X}_i}{\sum_i \tilde{X}_i}\right) \,.\, \log \text{MRPX}_i\right) - \log \text{MRPX}_i\right] \tag{16}$$

where $\tilde{X}_i = \left(\frac{X_i}{\theta_i^X - 1}\right)$. Define $w_i^X = \left(\frac{\tilde{X}_i}{\sum_i \tilde{X}_i}\right)$, and imputing the reallocation policy 16 back in expressions 15 and 13 provides with the following expression for reallocation gains under endogenous markups:

$$\text{APG-R}(\Gamma_i) = \sum_i \lambda_i \, \Gamma_i \left[\log \text{MRPX}_i - \left(\sum_i w_i^X \,.\, \log \text{MRPX}_i\right)\right] \tag{17}$$

[38]Notice that the knowledge of pass-through rate is necessary and sufficient to assess the reallocation gains under variable markups. I do not need information on how market structure or demand faced by firms will change in response to targeted subsidies. Indeed, a combination of those responses is exactly what the pass-through rates will capture.

It is clear from equation 17 that under variable markups, potential gains from reallocation will be affected by the pass-through rate. Under well-known case of monopolistic competition and CES demand, pass-through is complete ($\Gamma_i = 1$), and therefore, any reallocation targeted at exogenous wedges will increase aggregate productivity as intended. For example, in Hsieh and Klenow (2009), the expression 17 provides the exact quantification on productivity losses from misallocation when $\Gamma_i = 1$. Under variable markups, however, the pass-through rate is incomplete ($\Gamma_i < 1$). While the larger firms face more distortions and would need a large subsidy, they have lower pass-through rate and will change their prices less relative to their subsidies. This lowers gains from any targeted reallocation policy.

Pass-through and the demand curvature. The final task requires to decompose the gains in reallocation due to demand- and supply-factors. I do it in a parsimonious way and rely on the functional form of pass-through rate proposed in Weyl and Fabinger (2013) and Atkin and Donaldson (2015). I use the following general expression for Γ_i:

$$\Gamma_i = \left[1 + \frac{1+\delta_i}{\Phi_i}\right]^{-1} \frac{1}{\mu(\Phi_i, \sigma_i)} \equiv \left[1 + \frac{1}{\Phi_i} \cdot \frac{\chi_i}{\sigma_i}\right]^{-1} \frac{1}{\mu(\Phi_i, \sigma_i)} \tag{18}$$

where $\delta_i = \left[\frac{\partial \log\left(\frac{\partial P_i}{\partial Q_i}\right)}{\partial \log Q_i}\right]$ is the elasticity of the slope of inverse demand curve, and $\mu(\Phi_i, \sigma_i)$ is the markup from equation 1. With some rearranging, $\delta_i = \frac{\chi_i}{\sigma_i} - 1$, where $\chi_i = \left(1 + \frac{\partial \log\left(\frac{\partial Q_i}{\partial P_i}\right)}{\partial \log P_i}\right)$ is the elasticity of slope of demand ("super-elasticity") and σ_i is the demand elasticity. Therefore, the level of pass-through Γ_i depends on (i) elasticity of slope of demand χ_i, (ii) demand elasticity σ_i, and (iii) competitive structure of industry Φ_i. To assess how demand- and supply-factors affect the reallocation gains in 17 through their effects on pass-through rates, one would need to separately observe χ_i, σ_i and Φ_i. However, a primary challenge is that none of these parameters are observable to researchers — indeed, if they were observed one could have used that directly to compute pass-through rates. I next provide a strategy to separate out demand factors from competitive factors from firms' estimated markups and pass-through rate.

5.2 Identification of Γ, Φ, σ, and χ

I now describe the methodology to estimate firm-level pass-through rate Γ_i, and its underlying components.

Step 1: Estimate firm-level pass-through rates Γ_i: Firm-level pass-through Γ_i can be estimated using the information on prices, marginal costs and the following relationship:

$$\log P_{ijt} = \Gamma_i \log MC_{ijt} + \alpha_{ij} + \alpha_{jt} + \xi_{ijt} \tag{19}$$

I compute the pass-through rates both by using marginal costs directly and by instrumenting it with estimated quantity productivity (TFPQ). A potential issue in using marginal costs is that it is calculated using prices and markups, and therefore measurement error could generate upward bias in the estimates of Γ_i. Instrumenting the marginal costs with TFPQ addresses this issue. I also show later that the OLS and IV estimates of pass-through are not significantly different from one another, suggesting that the bias is not a primary concern.

Step 2: Recover estimates of competitive index Φ_i for homogeneous sector: Let $\hat{\mu}_i$ denote the estimate of firm markups μ_i derived in Section 3. Assumption 6 implies that any variation in markups in homogeneous sector arises only due to differences in the competitive index for firms (embedded in Φ_i). I can then use equation 1 to obtain the following relationship between markups, the slope of demand and the conduct parameter:

$$-\log\left(\hat{\mu}_{ij}^{-1} - 1\right) = \log \sigma^{\text{non-diff}} + \log \Phi_{ij} \tag{20}$$

where, following Atkeson and Burstein (2008) and Edmond, Midrigan, and Xu (2015), there is one-to-one mapping between firms' competitiveness and relative size of the firm within its industry.[39] This relationship is captured using the following polynomial model:[40] $\log \Phi_{ij} = \boldsymbol{\zeta}_z f(\log z_i) + \alpha_j + \epsilon_{ij}^{\phi}$. Substituting in equation 20 gives the following equation that I take to the data:

$$-\log\left(\hat{\mu}_{ij}^{-1} - 1\right) = \log \sigma^{\text{non-diff}} + \boldsymbol{\zeta}_z f(\tilde{\boldsymbol{z}}_{\boldsymbol{i}}) + \alpha_j + \epsilon_{ij}$$

The first term is identified through the constant in the regression. The second term captures the supply-side pass-through variation. The third term ensures that we compare firms within the same industry. The error term ϵ_{ij} captures the variation in Φ_i that is orthogonal to firms' market shares.

Step 3: Recover estimates of slope of demand: I use the estimates for $\hat{\Phi}_i$ obtained in Step 2 to estimate σ_i for firms in quality-differentiated sector using:

$$\hat{\sigma}_i = \left[\left(\frac{1}{\hat{\mu}_i} - 1\right)\hat{\Phi}_i\right]^{-1} \tag{21}$$

[39] The firms' demand elasticity in Atkeson and Burstein (2008) is a linear function of its market shares.

[40] I allow Φ_i to be a flexible function of firm's market share, which I proxy by employment: Specifically, I use the following relationship $\log \Phi_i^{-1} = \zeta_z^1 \log z_i + \zeta_z^2 (\log z_i)^2 + \nu_i$, where z_i is firm i's relative employment in its industry. The estimates are $\zeta_z^1 = -0.12$ (t-stat of -8.5 with errors clustered at firm-level), and $\zeta_z^2 = 0.028$ (t-stat of 4.19).

I assume that the relationship between firm-size and competitiveness index in Step 2, i.e. $\hat{\Phi}_i = \hat{\zeta}_{\mathbf{z}} f(\log z_i)$, also follows in the differentiated sector. This allows me to estimate $\hat{\Phi}_i$ in the differentiated sector. Combining $\hat{\Phi}_i$ with estimated markups 21 provides the estimates for slope of demand in differentiated sector. The slope of demand in homogeneous sector is $\sigma^{\text{non-diff}}$.

Step 4: Recover estimates of super-elasticity for all firms: Finally, I use the estimates for $(\hat{\Phi}_i, \hat{\sigma}_i)$ obtained in Step 2 and 3 to estimate χ_i for all firms. Let $\hat{\Gamma}_i$ denote the unbiased estimate of Γ_i obtained from from Step 1. I use the relationship 18 to obtain estimates of χ_i:

$$\hat{\chi}_i = \left(\frac{1}{\hat{\Gamma}_i \hat{\mu}_i} - 1 \right) \hat{\Phi}_i \, \hat{\sigma}_i \tag{22}$$

5.3 Results

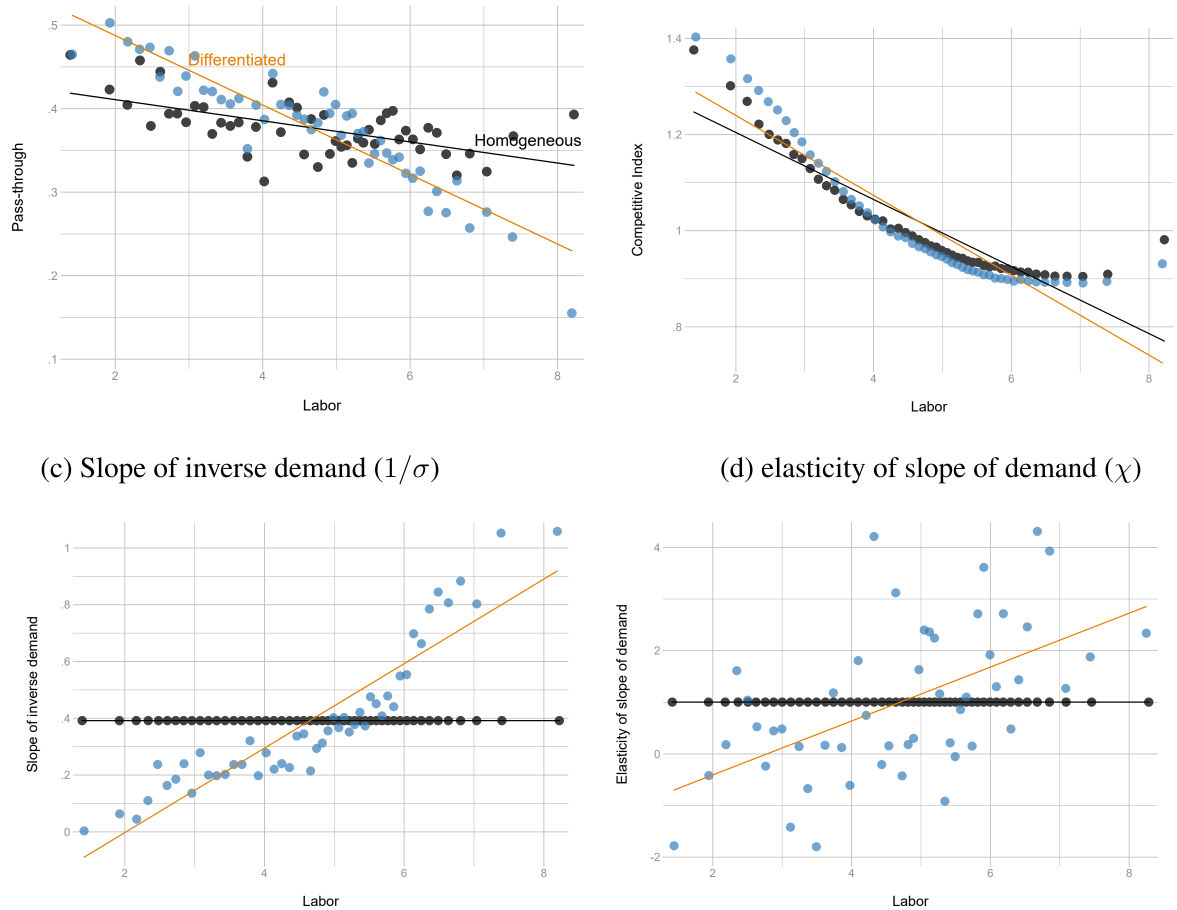
I start by documenting the results on pass-through rates, and its underlying components using the methodology described in Section 5.2. I then use the estimated parameters to quantify the aggregate productivity gains from reallocation under variable markups observed in the data, and under counterfactual scenarios with different parameter values of conduct and demand.

Pass-through rates. Table E.11 shows the results on pass-through estimates from equation 19. The average pass-through rate is 55 percent (OLS estimates, Column 1), and 70 percent (IV estimates, Column 4). Columns 2 and 5 show that larger firms pass-through less of changes in costs into their prices. Column 3 and 6 show that the negative relation between pass-through and firm size is stronger in quality-differentiated sectors.

Figure V shows the estimates of pass-through rate and its underlying components across the firm-size distribution separately for homogeneous and differentiated sectors. I winsorize all estimates at 5 percent. Panel (a) shows that pass-through rates are decreasing in firm size, and this relationship is stronger in quality differentiated sectors. Panel (b) shows the negative relationship between pass-through rate and firm size also reflects lower competition faced by larger firms. The similarity in the slope of two lines is just mechanical — by construction (Step 3 above) the relation between firm-size and competitiveness index is same across two sectors. Panel (c) shows that larger firms in quality-differentiated sector face higher slope of inverse demand (i.e., less elastic demand curve) relative to smaller firms. As demand composition is not a feature of homogeneous sector, the slope of inverse demand with firm-size is zero for this sector. Finally, Panel (d) shows that the elasticity of slope of demand is increasing in firm-size in the differentiated sector.

Counterfactuals. Next, I use these parameters to estimate aggregate productivity under various scenarios, and assess the role of demand factors in reducing aggregate productivity gains from reallocation. Column 2 of Table IV presents the results under planners' objective to equate MRP

Figure V: Pass-through rate, competitiveness, demand slope and curvature across firm size



(a) Pass-through rate (Γ)

(b) Competitive index (Φ)

(c) Slope of inverse demand ($1/\sigma$)

(d) elasticity of slope of demand (χ)

The figure shows the estimates of pass-through Γ_i (Panel (a)), competitiveness index Φ_i (Panel (b)) , slope of inverse demand $1/\sigma_i$ (Panel (c)), elasticity of slope of demand (super-elasticity) χ_i (Panel (d)) as a function of firm-size.

of material inputs across firms within a sector. I start by calculating gains from reallocation under the natural benchmark scenario of exogenous markups (Scenario (1)). When markups are assumed to be exogenous wedge, they do not react to underlying environment or policy changes. I plug in $\Gamma_i = 1$ along with the estimates of tax-subsidy policy, weighted-average MRPs and Domar weights in equation 17. The first row shows that assuming markups to be exogenous would give us an estimate of about 47 percent for the aggregate productivity gains from reallocation.

In the second row of Table IV, I allow for firms to adjust their markups in response to the tax-subsidy policy (Scenario (2)). As described in Section 5.1, this markup adjustment is captured by firm-level pass-through rates Γ_i. I plug in the estimates of Γ_i from the data in equation 17. As reported in the second row, while the estimated productivity gains from reallocation are still significant and positive (15.3 percent), they are an order of magnitude lower than the benchmark

Table IV: % change in aggregate productivity from reallocation

	Pass-through Γ_i considered:	**Reallocation gains from equating ....**	
		MRP inputs	markups
	(1)	(2)	(3)
(1) Complete pass-through	1	46.9%	36.6%
(2) Incomplete pass-through	actual (from data)	15.3%	8.8%
(3) Maximum competitiveness in data + estimated demand	$\Gamma(\chi_i, \sigma_i, \Phi^{\max})$	33.2%	57.0%
(4) Minimum competitiveness in data + estimated demand	$\Gamma(\chi_i, \sigma_i, \Phi^{\min})$	16.9%	16.8%

The table reports gains from a reallocation policy that equates marginal revenue products for materials across firms within 4-digit industries (Column 2), and markups (Column 3). The reallocation gains are calculated by averaging annual gains across the sample period.

case of complete pass-through. This is because high markup firms are also the firms that have the lowest pass-through of subsidies into their prices.

Next, in Scenario (3), I analyze the role of demand-based markup channel by applying the restriction that all firms face maximum competitiveness, while holding fixed their estimated slope of demand and super-elasticity. Specifically, I allow all firms to face the maximum competition within 4-digit industries every year observed in the data ($\Phi_i = \Phi^{\max}$). I use the estimated $\Phi^{\max}$ and plug it into:

$$\text{APG-R}\left(\chi_i, \sigma_i, \Phi_i^{\max}\right) = \sum_i \lambda_i \, \Gamma_i\left(\chi_i, \sigma_i, \Phi_i^{\max}\right)\left[\log \text{MRPX}_i - \sum w_i^X . \log \text{MRPX}_i\right]$$

The estimate in third row of Table IV shows that if all firms faced maximum competitiveness within the sector, aggregate productivity gains are 33.2 percent. While these gains are 13.7 percent point lower than the gains when pass-through is assumed to be completed, they are still order of magnitude higher than the observed pass-through in the data. This implies that the demand-driven markup dispersion lowers the aggregate productivity gains from reallocation by about 30 percent.[41]

Finally, I also created counterfactual Scenario (4) where I allow competitiveness index Φ_i to be the least competitive environment (that is, the minimum competition within 4-digit industries in the data $\Phi_i = \Phi^{\min}$)), while keeping fixed the estimated demand parameters. The results are reported

[41] This number is calculated by considering the amount of reduction in reallocation gains (46.9 percent to 33.2 percent) that can be explained by going from Scenario (1) of complete pass-through to Scenario (3) which uses estimated demand but holds competition to its maximum level estimated in the data.

in fourth row in Table IV. The aggregate reallocation gains are 16.9 percent which are closer to Scenario (2) that uses observed pass-through rates, suggesting that limited competition faced by firms also generates large misallocation losses. Increasing competition, therefore, can potentially reduce misallocation losses by as much as 50 percent, suggesting that markup variation due to supply-side factors is also an important source of allocative inefficiency (Edmond, Midrigan, and Xu 2015). In Column 3 of the table, I conduct the exercise with the objective function of equalizing markups (instead of MRP of inputs) and find similar results.

5.4 Caveats

The exercise above is a partial equilibrium analysis, and as such, comes with few caveats. First, it does not take into account entry and exits of firms. In this sense, the exercise is static in nature. This allows me to concentrate attention on *static misallocation* — and its implication for aggregate productivity — in a spirit that is closer to much of the existing work. An additional margin through which variable markups reduce aggregate productivity is the selection of firms, as suggested by Dhingra and Morrow (2019). Second, the first-order decomposition might not be a good approximation to quantify gains if the distortions are large and a subsidy can have higher-order effects on productivity growth. However, when trying to understand the impact of enacted policies under endogenous markups, one could sum over first-order approximations of policy effects each year to obtain the non-linear approximation of the effects of policy over a longer time horizon (Baqaee and Farhi 2019a). Third, my methodology does not explicitly allow for factor-biased technological change. Recent estimation methods have considered the role labor-augmenting technological change (e.g, Raval 2023). However, these methods do not allow for a generalized production function. Lastly, the analysis does not consider effects on consumer welfare. It has instead focused on the aggregate productivity in the manufacturing sector, given the available data. An assessment of consumer welfare would study the changes to consumer price indices across the income distribution. Such an exercise will require complete information on quality-adjusted product prices in the consumption baskets, which is not available in the NSS data. The methodology recently proposed by Atkin, Faber, Fally, and Gonzalez-Navarro (2020) provides a promising direction to estimate consumer welfare across the income distribution in absence of detailed price information. I leave the investigation of these themes for future work.

6 Conclusion

There is now an increasing evidence documenting higher markups for larger firms. The empirical evidence on the sources driving this correlation is, however, rather scarce. In this paper, I provide evidence on how demand-side characteristics affect the equilibrium distribution of markups across firms. I also assess the implications of this demand-driven markup dispersion for understanding misallocation losses. My results provide strong support for models that feature heterogeneous demand elasticities across firms, which are able to generate variable markups. However, I go a step further in documenting the interaction of consumer and firm heterogeneity in driving these variable markups. I show that heterogeneity in consumer preferences — that is, differences in their demand elasticities and preferences over quality — across income distribution translates into heterogeneity in markups charged by firms: lower demand elasticity of wealthier households allows larger firms to charge higher markups. While this demand-driven variable markups generate misallocation across firms, the losses from such misallocation are limited.

I use a sufficient statistic, firms' pass-through rate, to correct bias in aggregate reallocation gains under endogenous markup adjustments by firms. I find that pass-through rates are decreasing in firms size, with the relationship stronger in quality-differentiated sector. These differences in pass-through rates are driven by both differences in the slope of demand curve and market structure faced by firms. I propose a methodology — supported by the empirical evidence presented — that uses differences in markups and pass-through rates across homogeneous and differentiated sector to identify how differences in the demand characteristics across firms affect their pass-through rates. The main finding is that gains from reallocation are lower by about 30 percent under demand-driven variable markups than when markups are assumed to be exogenous.

Like much of data available in developing countries, I do not directly observe the characteristics of consumers that buy from firms. Yet this paper shows that inferences on how consumer demand affects firms' prices — and its underlying components — can still be made by combining available production data for firms with natural experiments. With separate data on prices and quantities (rather than revenues), — and despite imposing minimal assumptions on demand or market structure faced by firms — differences in markups and marginal costs across firms and sectors, and how firms change their prices in response to changes in their costs can inform us to a great extent about sources behind firms' market power and how that affects aggregate productivity.

References

Allen, T., and D. Atkin (2016): "Volatility and the Gains from Trade," *Working Paper*.

Asher, S., and P. Novosad (2012): "The Impact of Agricultural Output on Local Economic Activity India," *Working Paper*.

Asker, J., A. Collard-Wexler, and J. De Loecker (2014): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122(5), 1013–1063.

Atkeson, A., and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98(5), 1998–2031.

Atkin, D., A. Chaudhry, S. Chaudhry, A. K. Khandelwal, and E. Verhoogen (2015): "Markup and Cost Dispersion across Firms: Direct Evidence from Producer Surveys in Pakistan," *American Economic Review*, 105(5), 537–544.

Atkin, D., and D. Donaldson (2015): "Who's Getting Globalized? The Size and Implications of Intranational Trade Costs," *Working Paper*.

Atkin, D., B. Faber, T. Fally, and M. Gonzalez-Navarro (2020): "Measuring Welfare and Inequality with Incomplete Price Information," *Working Paper*.

Auer, R., A. Burstein, S. Lein, and J. Vogel (2023): "Unequal Expenditure Switching: Evidence from Switzerland," *The Review of Economic Studies*, p. rdad098.

Baldwin, R., and J. Harrigan (2011): "Zeros, Quality, and Space: Trade Theory and Trade Evidence," *American Economic Journal: Microeconomics*, 3(2), 60–88.

Baqaee, D. R., and E. Farhi (2019a): "A Short Note on Aggregating Productivity," *Working Paper*.

——— (2019b): "Productivity and Misallocation in General Equilibrium," *The Quarterly Journal of Economics*, 135(1), 105–163.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013): "Cross-Country Differences in Productivity: The Role of Allocation and Selection," *American Economic Review*, 103(1), 305–34.

Bils, M., P. J. Klenow, and C. Ruane (2018): "Misallocation or Mismeasurement?," *Working Paper*.

Brooks, W. J., J. P. Kaboski, Y. A. Li, and W. Qian (2021): "Exploitation of labor? Classical monopsony power and labor's share," *Journal of Development Economics*, 150, 102627.

Colmer, J. (2021): "Temperature, Labor Reallocation, and Industrial Production: Evidence from India," *American Economic Journal: Applied Economics*, 13(4), 101–24.

David, J. M., and V. Venkateswaran (2019): "The Sources of Capital Misallocation," *American Economic Review*, 109(7), 2531–67.

De Loecker, J., and P. K. Goldberg (2014): "Firm Performance in a Global Market," *Annual Review of Economics*, 6(1), 201–227.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, Markups, and Trade Reform," *Econometrica*, 84(2), 445–510.

de Roux, N., M. Eslava, S. Franco, and E. Verhoogen (2020): "Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments," *Working Paper*.

Dhingra, S., and J. Morrow (2019): "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," *Journal of Political Economy*, 127(1), 196–232.

Dixit, A. (1985): "Chapter 6 Tax policy in open economies," vol. 1 of *Handbook of Public Economics*, pp. 313–374. Elsevier.

Domar, E. D. (1961): "On the Measurement of Technological Change," *The Economic Journal*, 71(284), 709–729.

Edmond, C., V. Midrigan, and D. Y. Xu (2015): "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105(10), 3183–3221.

——— (2019): "How Costly Are Markups?," *Working Paper*.

Eslava, M., J. Haltiwanger, and N. Urdaneta (2023): "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand, and Wedges," *The Review of Economic Studies*.

Faber, B., and T. Fally (2020): "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," *Working Paper*.

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

Ganapati, S., J. S. Shapiro, and R. Walker (2020): "Energy Cost Pass-Through in US Manufacturing: Estimates and Implications for Carbon Taxes," *American Economic Journal: Applied Economics*, 12(2), 303–42.

Gandhi, A., and A. Nevo (2021): "Chapter 2 - Empirical models of demand and supply in differentiated products industries," vol. 4 of *Handbook of Industrial Organization*, pp. 63–139.

Goldberg, P. K., and R. Hellerstein (2012): "A Structural Approach to Identifying the Sources of Local Currency Price Stability," *The Review of Economic Studies*, 80(1), 175–210.

Gollin, D., and C. Udry (2021): "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture," *Journal of Political Economy*, 129(1), 1–80.

Gruber, J. (1997): "The Consumption Smoothing Benefits of Unemployment Insurance," *The American Economic Review*, 87(1), 192–205.

Haltiwanger, J., R. Kulick, and C. Syverson (2018): "Misallocation Measures: The Distortion That Ate the Residual," *Working Paper*.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Jayachandran, S. (2006): "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," *Journal of Political Economy*, 114(3), 538–575.

Johnson, R. (2012): "Trade and prices with heterogeneous firms," *Journal of International Economics*, 86(1), 43–56.

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 27(2), 1241–1277.

Kugler, M., and E. Verhoogen (2011): "Prices, Plant Size, and Product Quality," *The Review of Economic Studies*, 79(1), 307–339.

Leemput, E. V. (2016): “A passage to india: Quantifying internal and external barriers to trade.,” .

Li, N. (2021): “An Engel Curve for Variety,” *The Review of Economics and Statistics*, 103(1), 72–87.

Linder, S. B. (1961): *An Essay on Trade and Transformation*. New York: Wiley & Sons.

Mahajan, V., and R. K. Gupta (2011): *Non Farm Opportunities for Smallholder Agriculture*.

Martin, L. A., S. Nataraj, and A. E. Harrison (2017): “In with the Big, Out with the Small: Removing Small-Scale Reservations in India,” *American Economic Review*, 107(2), 354–86.

Mian, A., and A. Sufi (2014): “What Explains the 2007-2009 Drop in Employment?,” *Econometrica*, 82(6), 2197–2223.

Nakamura, E., and D. Zerom (2010): “Accounting for Incomplete Pass-Through,” *The Review of Economic Studies*, 77(3), 1192–1230.

Peters, M. (2020): “Heterogeneous Markups, Growth, and Endogenous Misallocation,” *Econometrica*, 88(5), 2037–2073.

Petrin, A., and J. Levinsohn (2012): “Measuring aggregate productivity growth using plant-level data,” *The RAND Journal of Economics*, 43(4), 705–725.

Rauch, J. E. (1999): “Networks versus markets in international trade,” *Journal of International Economics*, 48(1), 7 – 35.

Raval, D. (2023): “Testing the Production Approach to Markup Estimation,” *The Review of Economic Studies*, p. rdad002.

Restuccia, D., and R. Rogerson (2008): “Policy distortions and aggregate productivity with heterogeneous establishments,” *Review of Economic Dynamics*, 11(4), 707 – 720.

——— (2017): “The Causes and Costs of Misallocation,” *Journal of Economic Perspectives*, 31(3), 151–74.

Rodrik, D. (1987): “Policy Targeting with Endogenous Distortions: Theory of Optimum Subsidy Revisited,” *The Quarterly Journal of Economics*, 102(4), 903–911.

Rotemberg, M. (2019): “Equilibrium Effects of Firm Subsidies,” *American Economic Review*, 109(10), 3475–3513.

Rotemberg, M., and T. K. White (2020): “Measuring Cross-Country Differences in Misallocation,” *Working Paper*.

Santangelo, G. (2019): “Firms and Farms: The Local Effects of Farm Income on Firms’ Demand,” *Working Paper*.

Weyl, E. G., and M. Fabinger (2013): “Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition,” *Journal of Political Economy*, 121(3), 528–583.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012): “Monopolistic Competition: Beyond the Constant Elasticity of Substitution,” *Econometrica*, 80(6), 2765–2784.

Online Appendix for "Demand For Quality, Variable Markups and Misallocation: Evidence from India"

A A Simple Theoretical Framework (For Online Publication)

This section develops a simple model that features quality choice in a setting with heterogeneous households in consumption and heterogeneous firms in production. The model serves two purposes. First, it generates predictions on relation between firm-size, markups and costs that are consistent with the empirical correlations documented in Section 3. Specifically, the model predicts that markup dispersion in quality differentiated sector is generated due to assortative matching between firms and households. Second, and more importantly, the model generates testable prediction for how firms across the size distribution should change their markups in response to demand shocks across the income distribution, that I test in Section 4.

A.1 Model

The demand side features consumers that have non-homothetic preferences: consumers with different income levels vary in their quality valuations and demand elasticities. Specifically, when faced with identical prices, rich and poor households allocate their consumption expenditure differently across the quality ladder. The production side is a reduced-form version of the quality-choice model of Kugler and Verhoogen (2011) and Hallak and Sivadasan (2013) that features endogenous output quality choice across heterogeneous firms.

Demand. Consumers are indexed by h. As in Handbury (2019), consumers spend their income across two sectors: manufactured goods M and an outside good I. Their preferences follow a two-tier utility where the upper-tier utility depends on the consumption of outside good I: $U_h = (U_M(I_h), I_h)$. Following Handbury (2019), I assume that the outside good I is normal (and I_h is therefore analogous to the income level).[42] By making consumption on manufactured goods to be a function of outside good consumption, I allow introduction of non-homotheticity in a reduced-form manner, similar to Faber and Fally (2020). Each household derives utility from a product variety produced by firm i. Each firm produces a unique variety of product, and therefore

[42] I choose electricity to be the normal good, given the homogeneity of the good and its availability in all households' consumption baskets. I find that household expenditure on electricity follows a significant log-linear relation with respect to household income (slope of 0.248 with standard error of 0.005 when errors are clustered at town-level).

i indexes both firms and products. Utility of household h from manufactured goods is defined by:

$$U_M(I_h) = \left[\sum_{i=1}^{I} (\zeta_i^{\nu_h} Q_{hi})^{\frac{\sigma_h - 1}{\sigma_h}}\right]^{\frac{\sigma_h}{\sigma_h - 1}} \quad \text{s.t.} \quad \sum_i P_i Q_{hi} \leq I_h$$

where each variety has a quality ζ_i, $\sigma_h > 1$ is households' demand elasticity, $\nu_h > 0$ is households' taste for quality. I assume that household utility from consuming better quality increases with their income level such that $\nu_1 < \nu_2$ if $I_1 < I_2$.[43] These preferences are common across households but non-homothetic as the utility from manufactured goods depends on income level I_h as well as households' taste for quality ν_h and demand elasticity σ_h. There are two advantages of working with this structure. First, I keep the price elasticity of demand to be constant within income groups but allow them to vary across income groups. Second, I impose no restriction on how price elasticity of demand depends on income and rather estimate it from data.

Proposition A.1. *Average quality of household consumption basket increases in quality valuation* ν_h.

Proof: Define $s_{hi} = \frac{Y_{hi}}{\sum_i Y_{hi}}$ as share of household expenditure on product i. Differentiating w.r.t. taste for quality ν_h yields:

$$\frac{ds_{hi}}{d\nu_h} = (\sigma_h - 1)s_{hi}(\log \zeta_i - \sum_i s_{hi} \log \zeta_i) \tag{A.1}$$

This implies that household's expenditure shares within product group increase in ν_h for products with above average quality, and decrease in ν_h for below average quality products. Therefore, households with lower quality evaluations ν_h allocate higher share of their consumption expenditure to products with lower quality. ■

Production. Each firm produces a single variety of product subject to a fixed cost F. The profit function for the firms is given by:

$$\pi_i = P_i Q_i - C'(Q_i)Q_i - F = \left(1 - \frac{1}{\mu_i}\right) Y_i - F$$

where P_i is the price of the product i, Q_i is quantity sold by firm, $C(Q_i)$ is the total cost and $Y_i = P_i.Q_i$ are the total sales made by the firm. I assume that marginal costs are increasing in firm's product quality. Specifically, following Kugler and Verhoogen (2011) and Hallak and Sivadasan (2013), the functional form for marginal costs is such that the total cost of firm is increasing in its

[43] In recent work, Comin, Lashkari, and Mestieri (2021) develop a framework that can rationalize increased willingness to pay for product quality with income levels.

quality and decreasing in productivity Ω_i and is given by $C(Q_i;\zeta_i,\Omega_i) = \frac{\zeta_i^\alpha Q_i}{\Omega_i} + k\zeta_i$, where $k > 0$ and $\alpha > 0$. Therefore, marginal cost for the firm is $C'(Q_i) = \frac{\zeta_i^\alpha}{\Omega_i}$, and is increasing in the quality of the product. Define the price index P_h faced by consumer group h as:

$$P_h = \left(\sum_{i=1}^{I} \left(\frac{P_i}{\zeta_i^{\nu_h}} \right)^{1-\sigma_h} \right)^{\frac{1}{1-\sigma_h}} \tag{A.2}$$

Let μ_i be the markups over marginal costs defined by $P_i = \mu_i C'(Q_i)$ gives the following expression for consumer demand curve:

$$Y_{hi} = \zeta_i^{(\sigma_h-1)(\nu_h-\alpha)} \, \mu_i^{1-\sigma_h} \, P_h^{-1} \, \Omega_i^{(\sigma_h-1)} \, I_h \tag{A.3}$$

Total sales made by firm i is given by $Y_i = \sum_i Y_{hi}$, where $Y_{hi} = P_i \, Q_{hi}$.

Firms' Optimization. In equilibrium consumers maximize utility. Firms take the consumers demand curve A.3 as given and choose their markup, and quality to maximize their profits, subject to free entry (zero profits). As all firms face same problem, I suppress subscript i for convenience:

$$\max_{\mu,\zeta} \;\; \pi(\mu) = \left(1 - \frac{1}{\mu}\right) \sum_h Y_h(\mu,\zeta,\Omega,I) - F$$

Proposition A.2. *Product quality of a firm is increasing in its sales.*

Proof: Optimal quality produced by firm is given by:

$$\zeta = \frac{1}{k} \left[\left(\frac{\tilde{\sigma} - 1}{\tilde{\sigma}} \right) \; Y \;\; (\hat{\nu} - \alpha) \right]$$

where $\hat{\nu} = \left[\frac{\sum_h (\sigma_h - 1) \; \nu_h \; Y_h}{\sum_h (\sigma_h - 1) \; Y_h} \right]$, and $\tilde{\sigma}$ is defined below. Therefore, product quality of the firm is increasing in its sales. Intuitively, this is because for two firms with the same consumer base, the larger firm would be more profitable for a given quality upgrade. ■

Proposition A.1 and A.2 imply a sorting on product quality among consumer income distribution and firm size distribution — wealthier households have larger share of their consumption expenditure from larger firms. I refer to this pattern as *assortative matching* on product quality.

Proposition A.3 (Assortative Matching). *The share of firms' sales made to households with higher quality valuation ν_h increases in product quality ζ.*

Proof: Define $\psi_h(\mu,\zeta,\Omega,I_h) = \frac{Y_h(\mu,\zeta,\Omega,I_h)}{\sum_h Y_h(\mu,\zeta,\Omega,I_h)}$ as the share of firm's sales made to the consumer group h. Differentiating w.r.t. taste for quality ν_h yields:

$$\frac{d\psi_h}{d\nu_h} = \psi_h(\sigma_h - \tilde{\sigma}) \log \zeta \tag{A.4}$$

Differentiating w.r.t. taste for quality ζ yields:

$$\frac{d\psi_h}{d\nu_h} = \psi_h \left[(\sigma_h - 1)\nu_h - \sum_h (\sigma_h - 1)\nu_h \psi_h \right] \zeta^{-1} \tag{A.5}$$

This implies that share of sales for firms with high ζ is increasing in household's quality valuation ν_h for households with above average quality valuations, and decrease in ν_h for below average quality valuation. Therefore, firms with higher than average quality have larger share of sales to household with higher quality valuations. ■

As marginal costs are increasing in the underlying product quality, assortative matching implies that larger firms have higher marginal costs and wealthier households pay more for the products they consume. This is consistent with the correlations documented in Section 3. Next, I use equation A.3 to arrive at an expression for *firm-level* markup:

$$\mu = \frac{\sum_h \sigma_h Y_h(\mu, \zeta, \Omega, I_h)}{\sum_h (\sigma_h - 1) Y_h(\mu, \zeta, \Omega, I_h)} = \frac{\tilde{\sigma}}{\tilde{\sigma} - 1} \tag{A.6}$$

where $\tilde{\sigma}$ is the average demand elasticity faced by firm and given by:

$$\tilde{\sigma} = \frac{\sum_h \sigma_h Y_h(\mu, \zeta, \Omega, I_h)}{\sum_h Y_h(\mu, \zeta, \Omega, I_h)} = \sum_h \sigma_h \psi_h(\mu, \zeta, \Omega, I_h)$$

Equation A.6 allows for a new source of markup variation across firms: firms face heterogeneous market demand curves depending on composition of income groups demanding their products. These differences in demand composition faced by firms are dictated by assortative matching on product quality, leading to larger firms facing lower demand elasticities and charging higher markups. I term this as the demand-based markup channel.

Prediction A.1 (Cross-section prediction). *Under assortative matching, decreasing consumer demand elasticities with income levels imply that firm markups are increasing in firm-size.*

Proof. See Appendix B.1.2.

A.2 Markups responses to demand shocks.

I now use the framework to generate empirical prediction on how markups should change across the firm size distribution in response to demand shocks to poorest households (i.e., the households with highest price elasticity). Let I_{pt} be income for the poorest consumer group. The elasticity of markups with respect to I_{pt} is:

Prediction A.2 (Time-series prediction). *Firms lower their markups in response to an increase in demand from poor households. Additionally, the markup response is convex with respect to the firm size.*

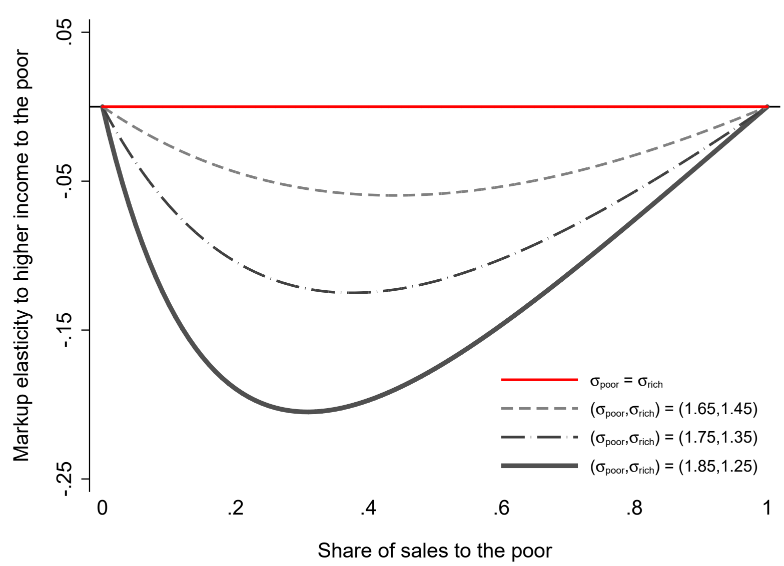
Proof. See Appendix B.1.3.

The following example illustrates this channel: Consider only two consumer groups in the economy - the poor and the rich consumers. As before, let I_{pt} be the income for the poor consumer group. Following equation B.9, the markup elasticities to the income shock I_{pt} is:

$$\frac{d \log \mu_{it}}{d \log I_{pt}} = -\frac{(\sigma_{poor} - \sigma_{rich})}{\tilde{\sigma}_{it}(\tilde{\sigma}_{it} - 1)} \times \psi_{poor,i,t} \times (1 - \psi_{poor,i,t})$$

Figure A.1 plots $\frac{d \log \mu_{it}}{d \log I_{pt}}$ from this specification as a function of share of sales made by firm to poorest income group, across various combinations of $(\sigma_{poor}, \sigma_{rich})$.

Figure A.1: Elasticity of markups to positive income shocks to poor (as function of share of sales made to the poor by firm)



The figure shows simulated relationship of elasticity of markups to positive income shocks to households that are more price-sensitive, as a function of share of sales made by the firm to those households.

Two findings emerge. First, the elasticity of markups is zero in absence of any heterogeneity in demand elasticities (i.e. under CES demand), and in absence of assortative matching. Second, under heterogeneous demand elasticities, markup elasticity is strictly convex with respect to share of sales made to the poor $\psi_{poor,i,t}$. The elasticity is highest for firms catering to both rich and poor households, while it approaches zero for firms making most of their sales to the poor households ($\psi_{poor,i,t} \to 1$), and for firms making most of their sales to the rich households ($\psi_{poor,i,t} \to 0$). The curvature of the function is also increasing in the demand elasticities gap between the two income groups. Intuitively, positive demand shocks to poor have a stronger positive effect on sales of firms that cater to a heterogeneous consumer base. This makes these firms pay more attention to the demand elasticity of its more price elastic consumer base, lowering their markups.

A.3 Alternative Demand System 1: Explicitly Additive Preferences

The model above has imposed few assumptions including a specific non-homothetic demand system. This particular demand system has the advantage of being simple, while providing tractable solutions and comparative statics. However, these functional forms are not crucial and in this section, I consider an alternate demand system with explicitly additive preferences. Consumers have directly explicitly additive preferences (Generalized Stone-Geary preferences) and have heterogeneous quality valuations. The production side is the same as Section A.

Demand. Consumer h has Stone-Geary preferences over the consumption goods Q_{hi}

$$U_h = \sum_i \left[\zeta_i^{\nu_h} \left(Q_{hi} - \underline{Q}_{hi} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

where, as before, ζ_i is the product quality; ν_h captures the consumer's valuation of quality, which I assume is strictly increasing in the exogenous income level I_h. The price elasticity of demand for consumer h for product i is given by:

$$\sigma_{hi} \equiv -\frac{P_i}{Q_{hi}} \frac{dQ_{hi}}{dP_i} = \sigma \left(1 - \frac{\underline{Q}_{hi}}{Q_{hi}} \right) \left(1 + \frac{P_i \underline{Q}_{hi}}{I_h - \sum_h P_i \underline{Q}_{hi}} \right)$$

The price elasticity of consumer is decreasing with the amount of residual income. Therefore, wealthier households have lower price elasticity of demand.

For firm i, the demand elasticity is the *sales-weighted* average of price elasticity of demand of its consumer base: $\tilde{\sigma}_i = \sum_h \sigma_{hi} \psi_{hi}$. The greater the firm's share of sales made to a particular income group ψ_{hi}, the higher is the weight the firm places on that group's price elasticity of demand σ_{hi}. Since larger firms make larger share of their sales to wealthier households — and because demand elasticity σ_{hi} is lower for wealthier households —, larger firms charge higher markups.

B Technical Appendix (For Online Publication)

B.1 Proof of propositions in Section 2

B.1.1 Proof of Proposition 1

I adopt the following assumptions: (i) the production function is Hicks-neutral, (ii) firms take input prices as given. These assumptions are also present in the cost-minimization exercise of markup estimation but I specify them here for this section to be self-contained. I make one additional assumption for exposition in this section: that the production function is homogeneous of degree one, that is, it exhibits constant return to scale (CRS).[44]

Under these assumptions, the production function takes the following general form: $Q = \Omega \,.\, F(\boldsymbol{X})$, where Q is firms' output (measured in physical quantities) and $\boldsymbol{X}$ is a vector of inputs. The first-order condition for a cost-minimizing firm with respect to inputs X^m and X^n with prices W^m and W^n is:

$$\frac{F_{X^m}(\boldsymbol{X})}{F_{X^n}(\boldsymbol{X})} = \psi\left(\frac{X^n}{X^m}\right) = \frac{W^m}{W^n} \tag{B.1}$$

where $\psi(.)$ is an increasing function. The first equality in the above expression following from Assumption (i) of homogeneity of the production function. Because all homogeneous functions are homothetic, $\frac{F_{X^m}(.)}{F_{X^n}(.)}$ is an increasing function of $\left(\frac{X^n}{X^m}\right)$. Defining $\Psi = \psi^{-1}(.)$, one can invert the second inequality to obtain the input demand function:

$$X^n = \Psi\left(\frac{W^m}{W^n}\right) X^m \tag{B.2}$$

Substituting the input demand in $Q = \Omega \,.\, F(\boldsymbol{X})$, and using the assumption of degree one homogeneity of production function provides:

$$\frac{Q}{\Omega} = X^m F(\boldsymbol{\gamma_i W}) \Rightarrow X^m(\boldsymbol{W}) = \frac{Q}{\Omega} g_m(\boldsymbol{W}) \tag{B.3}$$

where $g_m(.)$ is a function of input prices. Replacing the input demand in cost function $C(\boldsymbol{X}, \boldsymbol{W}) = \sum_m X^n W^m$ yields:

$$C(Q, \boldsymbol{W}) = \frac{Q}{\Omega} \sum_m g_m(\boldsymbol{W}) W^m \tag{B.4}$$

Define $\phi(\boldsymbol{W}) = \sum_m g_m(\boldsymbol{W}) W^n$. Finally, this provides us with the following functional form of marginal costs:

$$MC(\boldsymbol{W}) = \frac{\partial C(Q, \boldsymbol{W})}{\partial Q} = \frac{1}{\Omega}\phi(\boldsymbol{W}) \tag{B.5}$$

[44] Assumption (i) of CRS technology can be verified in the data and Table B.1 show that firms across most sectors indeed exhibit returns to scale of one.

Taking logs of equation B.5 shows that marginal costs (in logs) is additively separable into physical productivity (referred to as TFPQ), and a function of input prices $\phi(\boldsymbol{W})$. Taking partial derivative of marginal costs with respect to firm size yields following relationship:

$$\frac{\partial \log MC}{\partial \log \Omega} = -1 + \frac{\partial \log \phi(\boldsymbol{W})}{\partial \log \Omega} \tag{B.6}$$

In homogeneous sector firms face the same input prices, and the second term on the right hand side of equation is zero. Hence, $\frac{\partial \log MC}{\partial \log \Omega} = -1 < 0$ in homogeneous goods sector. For firms in differentiated sector, $\frac{\partial \log \phi(\boldsymbol{W})}{\partial \log \Omega} = \frac{\partial \log \phi(\boldsymbol{W})}{\partial \log \zeta}\frac{\partial \log \zeta}{\partial \log \Omega} > 0$. Hence, the relation between marginal costs and productivity depends on whether there are complementarities between the physical productivity and input quality (which is reflected positively in the input prices). Under Assumption 3 and 4 in main text, both terms on the right hand side of equation are positive. When these complementarities are large enough — for example in Kugler and Verhoogen (2011); Bastos, Silva, and Verhoogen (2018) — marginal costs can increase in firm productivity.[45] ■

B.1.2 Proof of Proposition 2

We are specifically interested in the relationship $\beta^{\mu} = \frac{\partial \log \mu}{\partial \log \Omega}$. Let $\beta^{u,h}$ and $\beta^{u,d}$ be the corresponding relationship in the homogeneous and differentiated goods sector, respectively. Taking the partial derivative of markup with respect to firm size yields:

$$\begin{aligned}
\beta^{u} = \frac{\partial \log \mu}{\partial \log \Omega} &= -\frac{\partial \log \left(1 - \frac{1}{\Phi}\frac{1}{\tilde{\sigma}}\right)}{\partial \log \Omega} \\
&= \left(\frac{1}{\Phi\tilde{\sigma}} - \frac{1}{\Phi\tilde{\sigma} - 1}\right)\frac{\partial \log \Phi\tilde{\sigma}}{\partial \log \Omega} \\
&= \frac{1}{1 - \Phi\tilde{\sigma}}\left[\frac{\partial \log \Phi}{\partial \log \Omega} + \frac{\partial \log \tilde{\sigma}}{\partial \log \Omega}\right]
\end{aligned}$$

As markups are positive and greater than 1, the first term on the right hand side is negative. In homogeneous sector firms face the same demand composition. Hence, the second term in the parenthesis is zero and $\beta^{u,h} = \frac{1}{1-\Phi\tilde{\sigma}} \cdot \frac{\partial \log \Phi}{\partial \log \Omega}$. If the conduct parameter Φ is decreasing in firm size, implying markups are increasing in firm-size in homogeneous sector ($\beta^{u,h} > 0$). In differentiated

[45] For general expression for correlation of marginal costs with firm size S, expression B.6 changes to $\frac{\partial \log MC}{\partial \log S} = -\frac{\partial \log \Omega}{\partial \log S} + \frac{\partial \log \phi(\boldsymbol{W})}{\partial \log S}$. Therefore, the same argument presented above applies whenever more productive firms are also larger in their size $\frac{\partial \log \Omega}{\partial \log S} > 0$. In the main analysis, I have used firms' labor force — instead of physical productivity Ω — as a proxy for firm size (S). The choice of labor force as a proxy for firm-size is intentional because unlike sales or productivity productivity (which is estimated through data), the use of labor force is not susceptible to measurement error in the independent variable that could be correlated with estimated marginal costs. Appendix Table E.2 shows that these results are robust if I use firms' total sales or fixed assets as alternate proxies for size.

sector, the second term in the parenthesis is negative because of assortative matching of wealthier less price-sensitive consumers to larger firms. To see this, let σ_{max} be the maximum demand elasticity, i.e., $\sigma_{\text{max}} = \max \{\sigma_h\}_{h=1}^{\mathcal{H}}$, and let the income group for which this happen be indexed by h^{max}. We can write the expersion for weighted demand elasticity as:

$$\tilde{\sigma} = \sigma_{\text{max}} - \sum_{h \neq h^{max}} (\sigma_{\text{max}} - \sigma_h)\, \psi_h \tag{B.7}$$

Taking partial with respect to firm produtivity Ω:

$$\frac{\partial \tilde{\sigma}}{\partial \Omega} = - \sum_{h \neq h^{max}} (\sigma_{\text{max}} - \sigma_h)\, \frac{\partial \psi_h}{\partial \Omega} < 0 \tag{B.8}$$

Thus the weighted elasticity $\tilde{\sigma}$ decreases in size if the share of sales made to less demand elastic income group increases in size. Hence, $\beta^{u,d} > \beta^{u,h}$ and the positive relationship of markups and firm size is stronger in differentiated sector than homogeneous goods sector. ■

B.1.3 Proof for Proposition 3

Let I_{pt} be income for the poorest consumer group. The elasticity of markups with respect to I_{pt} is:

$$\frac{d \log \mu_{it}}{d \log I_{pt}} = \frac{-1}{\tilde{\sigma}_{it}(\tilde{\sigma}_{it} - 1)} \frac{d\tilde{\sigma}_{it}}{d \log I_{pt}} = \frac{-1}{\tilde{\sigma}_{it}(\tilde{\sigma}_{it} - 1)} \sum_k \sigma_{kt} \psi_{kt} \frac{d \log \psi_{kt}}{d \log I_{pt}}$$

Solving and replacing for the last term in summation gives us:

$$\frac{d \log \mu_{it}}{d \log I_{pt}} = \frac{-\psi_{hi} \times (\sigma_p - \tilde{\sigma}_{it})}{\tilde{\sigma}_{it}(\tilde{\sigma}_{it} - 1)} = \frac{-\sum_{k \neq p} (\sigma_p - \sigma_k) \psi_{ki} \psi_{pi}}{\tilde{\sigma}_{it}(\tilde{\sigma}_{it} - 1)} \tag{B.9}$$

It is clear that markup responses to income shocks to the poor depends on (i) share of sales made by firm across income groups ψ_{ki}, and (ii) difference between demand elasticity of the poorest income group relative to other income groups $(\sigma_p - \sigma_k)$.

Define $\chi_{pit} \equiv \frac{d \log \mu_{it}}{d \log I_{pt}}$ as the elasticity of firm i's markup to income shocks to the poor in year t. When poorest households have highest price elasticity of demand (i.e. $\sigma_p > \sigma_k > 1 \;\, \forall \, k$), then equation B.9 implies $\chi_{pit} <= 0$. Thus, markups are weakly decreasing in response to positive income shocks to the poor.

Second, I analyze how χ_{pit} varies with share of firm i's sale made to the poor ψ_{pit}. I remove subscript i for convenience. Let $\chi_p(\psi_p) = f(\psi_p) \,.\, g(\psi_p)$, where

$$f(\psi_p) = \frac{-1}{\tilde{\sigma}(\tilde{\sigma} - 1)} < 0 \quad \text{and} \quad g(\psi_p) = \sum_{k \neq p} (\sigma_p - \sigma_k) \psi_k \psi_p > 0$$

To see that the function $\chi_p(\psi_p)$ has a unique minimum, we first solve for $\chi_p'(\psi_p)$

$$\chi_p'(\psi_p) \;=\; f(\psi_p)\,.\,\left[(f(\psi_p)(2\tilde{\sigma}-1)\,.\,\psi_p+1)\left(\sum_{k\neq p}(\sigma_p-\sigma_k)\psi_k\right)-\psi_p\,.\,\sum_{k\neq p}(\sigma_p-\sigma_k)\right]$$

Solving for $\chi_p'(\psi_p) = f'(\psi_p)\,.\,g'(\psi_p)$ = 0 gives

$$(f(\psi_p)(2\tilde{\sigma}-1)\,.\,\psi_p+1)\left(\sum_{k\neq p}(\sigma_p-\sigma_k)\psi_k\right)=\psi_p\,.\,\sum_{k\neq p}(\sigma_p-\sigma_k)$$

The left hand side is decreasing in ψ_p and the right hand side is increasing in ψ_p. Therefore, there exists a unique $\psi_p^* \in [0,1]$ for which $\chi_p'(\psi_p^*) = 0$. Next, we solve for $\chi_p''(\psi_p)$:

$$\chi_p''(\psi_p) = f''(\psi_p)\,.\,g(\psi_p) + 2\,.\,f'(\psi_p)\,.\,g'(\psi_p) + f(\psi_p)\,.\,g''(\psi_p) \tag{B.10}$$

$$\begin{aligned}
\text{where:}\quad f''g &= 2\frac{(f')^2}{f}g + 2f^2\left(\frac{d\tilde{\sigma}}{d\psi_p}\right)^2 g \quad \text{and} \quad f'g' = -\frac{(f')^2}{f}g \\
\text{and}\quad g''f &= -2\,.\,f\,.\,\left[\sum_{k\neq p}(\sigma_p-\sigma_k)\right]\,.\,>0
\end{aligned}$$

Substituting these expressions in B.10

$$\chi_p''(\psi_p) \;=\; 2f^2 - 2\,.\,f\,.\,\left[\sum_{k\neq p}(\sigma_p-\sigma_k)\right]\,.\,>0$$

Therefore, χ_{ht} is a convex function with a unique minimum. ■

B.2 Estimating markups and marginal costs

This section provides broad overview of the procedure for estimating markups in the Indian manufacturing data. The framework primarily builds on the methodology in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for estimating markups for multi-product firms.

B.2.1 Estimation.

Framework. Consider a production function for firm i and product j in year t:

$$Q_{ijt} = \Omega_{it} \; . \; F_{jt}(\boldsymbol{X}_{ijt}, \boldsymbol{K}_{ijt}) \tag{B.11}$$

where $\{\boldsymbol{X}_{ijt}, \boldsymbol{K}_{ijt}\}$ is a vector of input X and capital stock K which is assumed to be dynamic. Let the adjustment costs of any input $V \in \{\boldsymbol{X}, \boldsymbol{K}\}$ be captured by the function $\kappa(V_{it}, V_{it-1})$, and price schedules be given by W^{V}_{ijt}. We adopt the following set of assumptions to estimate markups:

Assumption B.1. *A firm engages in cost minimization taking output quantity at time t as given.*

Assumption B.1 is regularity condition for firm optimization. The associated Lagrangian function for any product j at time t is

$$\mathcal{L}(\boldsymbol{X}_{ijt}, \boldsymbol{K}_{ijt}, \lambda_{ijt}) = W^{\boldsymbol{X}}_{ijt}(\boldsymbol{X}_{ijt})\boldsymbol{X}_{ijt} + W^{\boldsymbol{K}}_{ijt}\boldsymbol{K}_{\text{ijt}} + \lambda_{ijt}(Q_{ijt} - Q_{ijt}(\boldsymbol{X}_{ijt}, \boldsymbol{K}_{ijt}, \Omega_{it})) \tag{B.12}$$

Next, there exists at least one input X for all firms that satisfies the following:

Assumption B.2. *There exists an input X with no adjustment costs, i.e., $\kappa(.,.) = 0$.*

Assumption B.3. *The input X is chosen statistically and is variable.*

Assumption B.4. *The firm is a price taker and exerts no monopsony power over the input X, i.e., $W^X(X_{it}) = W^X_{it}$.*

Assumption B.5. *Production $F(.)$ is continuous and twice differentiable in X.*

Assumption B.2 rules out inventories or adjustment costs. The presence of a variable and static input (Assumption B.3) implies that it is chosen in the same time period it is used and only affects current profits. This rule out dynamic considerations. Assumption B.4 implies that firms minimize cost taking input prices W^X_{ijt} at time t as given. Assumption B.5 ensures that inputs can be characterised through their first order conditions. The input that satisfies Assumptions B.2-B.5 is considered a flexible input. Next, I adopt the following assumption on firm's production technology:

Assumption B.6. *The production technology is product-specific.*

Assumption B.7. *Firm productivity Ω_{it} is Hicks-neutral, log-additive and firm-specific.*

Assumption B.8. *Expenditures on all variable and fixed inputs are attributable to products.*

Assumption B.6 implies that even though productivities across firms might differ, a single-product firm and a multi-product firm that produce the *same* product have same production technology $F(.)$. Assumption B.7 implies that a multi-product firm has the same productivity across all the product lines. Assumption B.8 ensures that the expenditure on input X is always attributable to production of output by firms.

Notice that the approach makes no assumption on the nature of demand, or the underlying nature of market competition, or returns to scale. The main challenge is to have the existence of a static input X that is free of adjustment costs and for which firms are price-takers. If there exists such an input, then the first-order condition from the cost-minimization problem B.12 for firms yields the following:

$$W_{ijt}^X = \lambda_{ijt} \frac{\partial Q_{ijt}(.)}{\partial X_{ijt}}$$

where $\lambda_{ijt} = \frac{\partial \mathcal{L}_{ijt}}{\partial Q_{ijt}}$ is the marginal cost of production. Rearranging terms above, multiplying both side with $\frac{X_{ijt}}{Q_{ijt}}$, and using $\mu_{ijt} = \frac{P_{ijt}}{\lambda_{ijt}}$ yields:

$$W_{ijt}^v \frac{X_{ijt}}{Q_{ijt}} = \frac{P_{ijt}}{\mu_{ijt}} \frac{\partial Q_{ijt}(.)}{\partial X_{ijt}} \frac{X_{ijt}}{Q_{ijt}}$$

This provides me with the main expression of markups at firm-product level:

$$\mu_{ijt} = \theta_{ijt}^X \left(\alpha_{ijt}^X\right)^{-1}, \quad \text{where} \quad \underbrace{\theta_{ijt}^X = \frac{\partial \log Q_{ijt}(.)}{\partial \log X_{ijt}}}_{\substack{\text{output elasticity} \\ \text{(obtained by estimating } F_{jt}(.))}}, \qquad \underbrace{\alpha_{ijt}^X = \frac{W_{ijt}^X X_{ijt}}{P_{ijt} Q_{ijt}}}_{\substack{\text{share of input expenditure} \\ \text{(only observe for single product firms)}}}$$

In addition to the fact that θ_{ijt}^X needs to be estimated, the main challenge when working with multi-product firms is that α_{ijt}^X is not observed in the data at the firm-product-level. I eventually estimate markup using:

$$\hat{\mu}_{ijt} = \hat{\theta}_{ijt}^X \cdot \frac{P_{ijt} Q_{ijt}}{\exp(\hat{\rho}_{ijt}) \tilde{W}_{it}^X} \tag{B.13}$$

where ρ_{ijt} is the share of input expenditure attributable to product j. The data provides information on $(P_{ijt}, Q_{ijt}, \tilde{W}_{it}^X)$, where $\tilde{W}_{it}^X$ is the total expenditure on flexible input X. The estimation of $\hat{\mu}_{ijt}$ involves estimation of production function parameter θ_{ijt}^X and the input allocation $\hat{\rho}_{ijt}$.

Production Function Estimation. Taking the logarithm of the production function gives:

$$q_{ijt} = f_{jt}(x_{ijt}, k_{ijt}; \boldsymbol{\theta}) + \omega_{it} \equiv f_{jt}(\boldsymbol{z}_{ijt}; \boldsymbol{\theta}) + \omega_{it} \tag{B.14}$$

where notation in small caps denote logarithms of corresponding large cap variables. Thus, any changes to output over time occurs due to either (i) changes in input quantities or (ii) unanticipated shocks to productivity. Here $\boldsymbol{z}_{ijt} = \{x_{ijt}, k_{ijt}\}$ is a vector of (log) physical inputs and $\omega_{it} = \log(\Omega_{it})$. The production coefficients $\boldsymbol{\theta}$ need to be identified.

Three biases arise in the estimation of the production function. First, output price-bias could arise when output is constructed by deflating firm revenues by an industry-level price index. A difference from existing work that relied on data on revenue is that here q_{ijt} is in physical units of output. This solves the output-price bias. Second, because I only observe input expenditure, and not input quantities, I need to modify the above expression with the use of input expenditure: $z_{ijt} = \rho_{ijt} + \tilde{z}_{it} - w^z_{ijt}$, using $W^Z_{ijt} Z_{ijt} = \tilde{\rho}_{ijt} \left[\sum_p W^Z_{ijt} Z_{ijt}\right] = \tilde{\rho}_{ijt} \tilde{Z}_{it}$. Here $\tilde{z}_{it}$ is the firm-level expenditure on input Z and w^Z_{ijt} is the deviation of the unobserved (log) firm-product-specific price from the (log) industry-wide input price index. This yields the following decomposition of expression B.14:

$$q_{ijt} = f_j(\tilde{\boldsymbol{z}}_{\boldsymbol{it}}; \boldsymbol{\theta}) + \underbrace{a(\rho_{ijt}, \tilde{\boldsymbol{z}}_{\boldsymbol{it}}, \boldsymbol{\beta})}_{\text{Input Allocation Bias}} + \underbrace{b(\boldsymbol{w}_{ijt}, \rho_{ijt}, \tilde{\boldsymbol{z}}_{\boldsymbol{it}}, \boldsymbol{\theta})}_{\text{Input Prices Bias}} + \omega_{it}$$

The objective is to address the two sources of biases: "Input Allocation Bias" and "Input Prices Bias". I now discuss the steps involved in addressing these biases, and the estimation of production function and input allocation.

Addressing input allocation bias. I address input allocation bias by focusing on single product firms. For these firms, $\rho_{ijt} = 1$ and hence $a(.) = 0$. That is for single-product firms, the true production function will not suffer from input allocation bias. I also drop sub-script j due to its redundancy for single-product firms:

$$q_{it} = f(\tilde{\boldsymbol{z}}_{\boldsymbol{it}}; \theta) + B(\boldsymbol{w}_{\boldsymbol{it}}, \tilde{\boldsymbol{z}}_{\boldsymbol{it}}, \boldsymbol{\theta}) + \omega_{it}$$

where $b(\boldsymbol{w}_{\boldsymbol{it}}, \tilde{\boldsymbol{x}}_{\boldsymbol{it}}, \boldsymbol{\theta})$ is the input-prices bias. I use three inputs in the (deflated) input expenditure vector $\tilde{z}_{it}$: labor ($\tilde{l}$), materials ($\tilde{m}$) and capital ($\tilde{k}$). Thus, $\tilde{z}_{it} = \{\tilde{l}_{it}, \tilde{m}_{it}, \tilde{k}_{it}\}$. I also address the selectivity of firms, that is the entry and exit of firms in and out of single-product firms using strategy in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016).

Addressing input price bias. I now address input-price bias. As I only see expenditure on inputs and not their quantities, the procedure requires treatment of unobserved input prices. I address this using the following procedure: I assume that input price function depends on firms location $\boldsymbol{G_i}$ and input quality v_{it}. Information on input quality can be obtained from output price p_{it}, market share $\boldsymbol{ms_{it}}$, product dummies $\boldsymbol{D_j}$, and location $\boldsymbol{G_i}$. The idea is that in absence of direct measures of input quality, information on output prices and their market share within a product category and location are informative of input quality (Kugler and Verhoogen 2011). I also include rain shocks in my estimation to allow for the possibility that rain shocks may, but need not, change input prices. I assume an input price control function: $w_{it}^x = w_t(p_{it}, \boldsymbol{ms_{it}}, \boldsymbol{D_j}, \boldsymbol{G_i}, r_{it})$. The input price bias function takes the form:

$$b(w_{it}, \tilde{z}_{it}, \boldsymbol{\theta}) = b((p_{it}, \boldsymbol{ms_{it}}, \boldsymbol{D_j}, \boldsymbol{G_i}, r_{it}) \times \boldsymbol{\tilde{z}_{it}^c}; \boldsymbol{\theta}, \boldsymbol{\delta})$$

where $\boldsymbol{\tilde{z}_{it}^c} = \{1, \boldsymbol{\tilde{z}_{it}}\}$. This allows for input expenditure vector to affect input prices by itself and separately through the interaction with the input price control function.

Productivity process, moment conditions and identification. Next, I next the identification of production function. I follow literature on production function estimation and control for unobserved productivity ω_{it} using static input demand equation for materials: $\tilde{m}_{it} = m_t(\Omega_{it}, \tilde{k}_{it}, \tilde{l}_{it}, \boldsymbol{\kappa}_{it})$, where $\boldsymbol{\kappa}_{it} = \{\boldsymbol{G_i}, r_{it}, p_{it}, \boldsymbol{D_j}, \boldsymbol{ms_{it}}\}$. Inverting this provides a control function for productivity: $\omega_{it} = h_t(\boldsymbol{\tilde{z}_{it}}, \boldsymbol{k_{it}})$. To estimate the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ I form moments based on innovation in productivity shock ξ_{it}:

$$\omega_{it} = g(\omega_{it-1}, r_{it-1}, SP_{it}) + \xi_{it}$$

where SP is the probability of remaining single-product. Again, I include local rain shocks r_{it} in the last year to allow for the possibility that it may affect productivity. Next, I estimate the production function parameters using the following steps standard in the literature (see De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) for detailed a discussion on the estimation techniques).

1. Use the original production equation with input price bias and run the first stage:

$$q_{it} = \phi_t(\boldsymbol{\tilde{z}_{it}}, \boldsymbol{\kappa_{it}}) + \epsilon_{it} \quad \text{where} \quad \phi_t(.) = f(\boldsymbol{\tilde{z}_{it}}; \boldsymbol{\theta}) + B(\boldsymbol{w_{it}}, \boldsymbol{\tilde{z}_{it}}, \boldsymbol{\theta}) + \omega_{it}$$

2. This allows to get productivity as a function of $(\boldsymbol{\theta}, \boldsymbol{\delta})$:[46]

$$\omega_{it}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \hat{\phi}_{it} - f(\boldsymbol{\tilde{z}_{it}}; \boldsymbol{\theta}) - b((p_{it}, \boldsymbol{ms_{it}}, \boldsymbol{D_j}, \boldsymbol{G_i}, r_{it}) \times \boldsymbol{\tilde{z}_{it}^c}; \boldsymbol{\theta}, \boldsymbol{\delta})$$

[46] I assume a translog functional for $f(.)$. Unlike Cobb-Douglas, the use of translog function has the advantage that the output elasticities with respect to inputs depend on the level of input factors. As input factors are observed in the data for each year, the use of translog functional form also allows for time-varying output elasticities with respect to each input. The qualitative results remain robust if Cobb-Douglas production function is used instead.

3. Obtain the innovation in productivity as function of $(\boldsymbol{\theta}, \boldsymbol{\delta})$ using the law of motion for ω_{it}:

$$\xi_{it}(\boldsymbol{\theta}, \boldsymbol{\delta}) = \omega_{it}(\boldsymbol{\theta}, \boldsymbol{\delta}) - E(\omega_{it}(\boldsymbol{\theta}, \boldsymbol{\delta})|\omega_{it-1}, r_{it-1}, SP_{it})$$

4. Finally, build moment conditions that identify the parameters using $E(\xi(\boldsymbol{\theta}, \boldsymbol{\delta})\boldsymbol{Y_{it}'}) = 0$ where $\boldsymbol{Y_{it}} = \{m_{it-1}, l_{it}, k_{it}, \boldsymbol{\kappa_{it-1}}\}$ along with the higher order terms and interactions

Recovering input allocation for multi-product firms. The final step requires estimation of input allocation parameter for multi-product firms, $\rho_{ijt} = \ln \frac{W_{ijt}^X X_{ijt}}{\tilde{X}_{it}} \forall X \in \{V\}$. To do so, I first eliminate unanticipated shocks and measurement error using $\hat{q}_{ijt} \equiv E[q_{ijt}|\phi_t(\tilde{z}_{it}, \kappa_{it})]$. The production function can then be written as: $\hat{q}_{ijt} = f(\boldsymbol{\tilde{z}_{it}}, \boldsymbol{\hat{\theta}}, \hat{w}_{ijt}, \rho_{ijt}) + \omega_{it}$, where $\hat{q}_{ijt}$ is obtained through first-stage estimation. I use translog for the production function functional form. I can use the estimation of $\hat{\theta}$ to decompose this translog into a component separately dependent on ρ_{ijt}:

$$\hat{\omega}_{ijt} \equiv \hat{q}_{ijt} - f_1(\boldsymbol{\tilde{z}_{it}}, \boldsymbol{\hat{\theta}}, \hat{w}_{ijt}) = f_2(\boldsymbol{\tilde{z}_{it}}, \hat{w}_{ijt}, \rho_{ijt}) + \omega_{it}$$

where I have $\hat{w}_{ijt}$ from the input price estimation. Using the translog functional form for the production function yields: $\hat{\omega}_{ijt} = \omega_{it} + \hat{a}_{ijt}\rho_{ijt} + \hat{b}_{ijt}\rho_{ijt}^2 + \hat{c}_{ijt}\rho_{ijt}^3$. This expression provides with $J+1$ equations in $J+1$ unknowns $(\omega_{it}, \rho_{i1t}, ..., \rho_{iJt})$ for each multi-product firm-year. Recall that all the parameters $(\hat{a}_{ijt}, \hat{b}_{ijt}, \hat{c}_{ijt})$ are functions of $(\hat{\theta}, \hat{w}_{ijt})$. With the estimates of ρ_{ijt}, the markups can be obtained from the expression B.13.

B.2.2 Results.

In this subsection, I first present summary statistics on output elasticities and markups, across sectors. I then cross-validate the measures of estimated markups and marginal costs by analyzing correlations in the data and comparing it to correlations documented in other settings.

Summary statistics. Table B.1 reports the output elasticities and returns to scale across industries and on average. The estimated coefficients for most sectors are close to constant returns to scale, with modest within-industry variation. The average returns to scale in the Indian manufacturing sector is 1.06.

Table B.2 reports the mean and median markups across each two digit industry in the manufacturing sector. The mean markup is 2.84 and the median markup is 1.50 with a standard deviation of 5.65, suggesting wide variation in markups across firms. These averages, however, mask considerable heterogeneity across industries. For example, the median markup about 1 in non-metallic minerals, whereas it is 2.42 for firms in industries that produce computing equipment.

Table B.1: Output Elasticities and Returns to Scale

	Labor	Capital	Material	RTS		Labor	Capital	Material	RTS
Sector	(1)	(2)	(3)	(4)	Sector	(1)	(2)	(3)	(4)
Food and beverages	0.06	0.05	0.82	0.94	Non-metal minerals	0.37	0.07	0.51	0.95
	[0.05]	[0.04]	[0.10]	[0.09]		[0.19]	[0.06]	[0.22]	[0.17]
Tobacco products	0.42	0.06	0.80	1.27	Basic Metals	0.06	0.04	0.92	1.02
	[0.19]	[0.05]	[0.14]	[0.21]		[0.05]	[0.03]	[0.06]	[0.06]
Textiles	0.19	0.03	0.96	1.17	Fabricated metal	0.26	0.16	0.74	1.17
	[0.12]	[0.03]	[0.05]	[0.15]		[0.19]	[0.09]	[0.19]	[0.15]
Wearing Apparel	0.35	0.04	0.52	0.91	Machinery	0.13	0.09	0.80	1.02
	[0.11]	[0.04]	[0.25]	[0.16]		[0.13]	[0.09]	[0.14]	[0.09]
Leather products	0.14	0.09	0.89	1.13	Electric	0.13	0.16	0.87	1.15
	[0.09]	[0.04]	[0.09]	[0.08]		[0.08]	[0.10]	[0.07]	[0.16]
Paper products	0.50	0.25	0.63	1.39	Motor vehicles	0.06	0.06	0.83	0.95
	[0.28]	[0.19]	[0.19]	[0.40]		[0.03]	[0.04]	[0.05]	[0.04]
Printing	0.15	0.14	0.71	1.00	Other transport	0.22	0.29	0.66	1.17
	[0.12]	[0.12]	[0.07]	[0.18]		[0.11]	[0.14]	[0.18]	[0.15]
Chemicals	0.06	0.02	0.89	0.97	Furniture	0.38	0.15	0.47	1.00
	[0.03]	[0.02]	[0.04]	[0.04]		[0.23]	[0.13]	[0.24]	[0.22]
Rubber and Plastic	0.17	0.29	0.40	0.87	**Average**	**0.21**	**0.12**	**0.73**	**1.06**
	[0.16]	[0.21]	[0.21]	[0.26]					

Notes: The table reports the estimated output elasticities for the production function, estimated within 2-digit industries. Columns 1-3 report the estimated output elasticity for each factor of production. Standard deviations of the output elasticities are reported in brackets. Column 4 reports the returns to scale.

Table B.2: Markups, by industry

	Markups			Markups	
Industry	Mean	Median	Industry	Mean	Median
Food and beverages	1.46	1.10	Non-metallic minerals	1.30	0.97
Tobacco products	2.53	2.37	Basic metals	2.52	1.80
Textiles	2.54	1.75	Fabricated metal	3.75	1.82
Clothing	3.14	1.08	Machinery	6.23	2.16
Leather products	4.15	1.93	Electrical mach. & comm.	3.87	1.76
Wood products	3.67	1.94	Medical equipments	5.83	2.42
Paper products	1.28	1.17	Automobiles	5.50	1.60
Printing and publishing	3.19	1.42	Other transportation	3.35	1.29
Chemicals	3.38	1.77	Furniture	2.66	1.50
Rubber and plastic	3.72	1.34	**Total**	**2.84**	**1.50**

Notes: The table displays the mean and median markups across 2-digit industries between 1998-2009. The tables trims observations that are below and above 5th and 95th percentile in each industry.

Cross-validation. Next, I perform three exercises to validate the estimates of markups and marginal costs. First, I examine how markups vary with firms' exporting behavior. There is extensive evidence that markups are systematically higher for exporting firms than domestic firms, and markups increase upon export entry (De Loecker and Warzynski 2012; Atkin, Khandelwal, and Osman 2017; Garcia-Marin and Voigtländer 2019). Although my sample size for exporters is small, I do find that markups are higher for exporters (Columns 1 -3 of Table B.3), and are increasing in share of sales exported by firms (Columns 4 - 6 of Table B.3).

Table B.3: Markups and export status

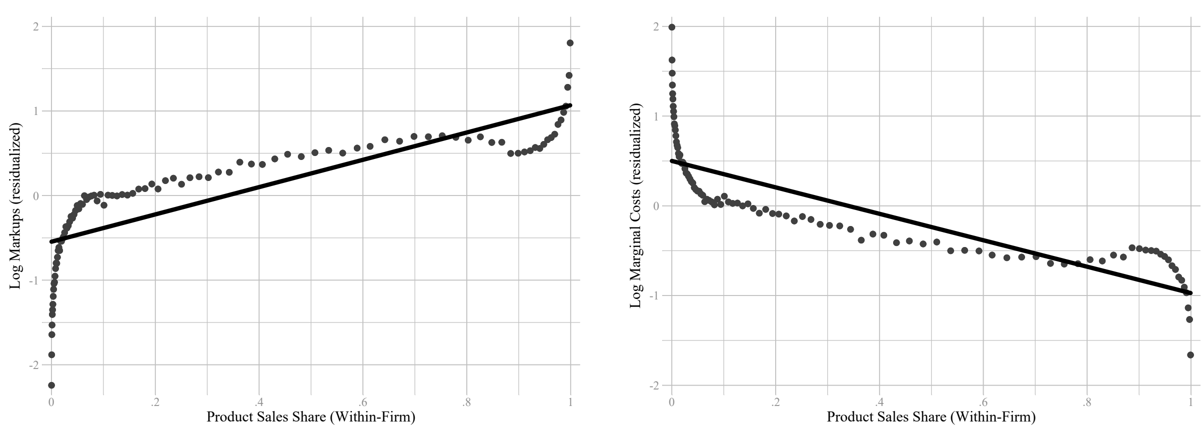
	Dependent variable: log (markup)					
	(1)	(2)	(3)	(4)	(5)	(6)
1(exporter)	0.076***	0.067***	0.060*	-	-	-
	[0.019]	[0.020]	[0.036]	-	-	-
% of sales exported	-	-	-	0.093***	0.091***	0.168**
	-	-	-	[0.033]	[0.034]	[0.073]
Firm-product f.e.			✓			✓
Product-year f.e.	✓	✓	✓	✓	✓	✓
District-year f.e.		✓	✓		✓	✓

Notes: Standard errors are clustered at firm-level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Second, mirroring De Loecker, Goldberg, Khandelwal, and Pavcnik (2016), I analyze how markups and marginal costs vary across products within a firm as function of their share of sales. Theoretical work by Mayer, Melitz, and Ottaviano (2014) suggests that multi-product firms feature a core competency wherein their core product has the lowest marginal cost. Figure B.1 provides evidence consistent with this hypothesis. It plots the markups and marginal costs against the share of sales made for that product within each firm. Markups rise as the firm move towards its core competency, and costs decrease. These correlations are obtained without imposing any assumptions on the demand system or market structure. Despite this, the patterns are remarkably consistent with the multi-product firm literature.

Figure B.1: Markups and marginal costs as share of sales within-firm

(a) Markups (b) Marginal Costs



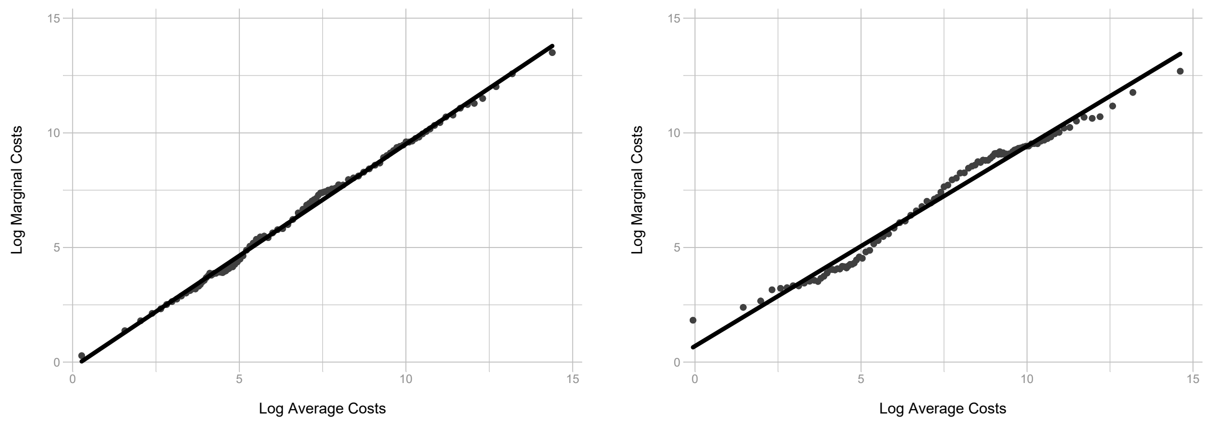
Notes: Markups and marginal costs are demeaned using product-year, firm-year and district-year fixed effects and outliers are trimmed at above and below 95th percent and 5th percent.

Third, Figure B.2 reports the correlations between marginal and average cost of production from ASI data, at the firm-product level. Panel (a) reports the correlation for single-product firms. In Panel (b), I also include multi-product firms in the sample. To calculate the average costs at firm-product level for multi-product firms, I multiply the firm-level cost reported in the ASI data with the share of sales across products within the firm. The figures show that marginal costs are tightly related to the average costs for the firms.

Together, these correlations provide credibility on the estimates of markups and marginal costs obtained from the cost-minimization approach.

Figure B.2: Relation between marginal costs and average costs

(a) Single product firms (b) All firms



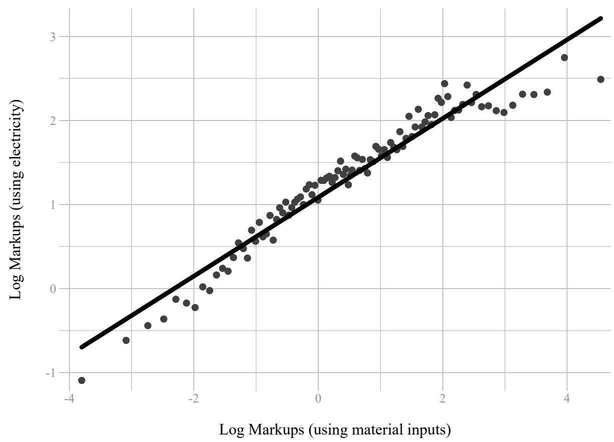
Notes: The figure plots (log) marginal costs and (log) average costs for single-product and for all firms.

B.2.3 Robustness.

Choice of flexible input. While the production function approach to identify markups has many advantages, it does come with few limitations. One key identification assumption is the presence of one input which is flexible. Identifying such an input is challenging, and I have followed the IO literature and considered material inputs as flexible. As an alternative, I consider electricity as the flexible input following the recommendation in Kim (2016). As shown in Figure B.3, the markups estimates using electricity as flexible input display a strong correlation to the markups estimated using material inputs.

The strong positive correlation between estimated markups from the two inputs is also reassuring for three other reasons. First, the electricity prices are regulated by state governments making it less likely for firms to exert monopsony power over this input. Second, given the sample on manufacturing firms, many of which are producing intermediate outputs that are inputs to other firms, markups for some firms could imply markdowns for firms using those products as

Figure B.3: Relation between markups estimated using different proxies for flexible input



Notes: The figure plots estimates of (log) markups using electricity (y-axis) and material inputs (x-axis).

inputs. Electricity as an input is not prone to this issue. Third, as describe above, output product differentiation generates differences in input quality, and hence, input prices (input price bias described above). However, electricity is a homogeneous input and is less susceptible to input price bias. Altogether, the strong positive association across the two estimated markups suggests that material inputs are a suitable proxy for flexible input in my setting.

Role of markdowns. Another assumption with the identification of markups is that the firms should be price takers in the flexible input. There is some recent evidence on monopsony in market for material inputs. Under monoposony power, the ratio of output elasticities and revenue share of that input identifies a "net markup", which is the product of output markup and input markdown. This happens because markdowns act as a "wedge" in the marginal costs across firms, increasing it for firms that can exert market power in the input market. If that were the case in my setting, then the estimated correlation between firm-size and marginal costs would be biased towards zero. To see this, I introduce a markdown wedge $\mathcal{M}_i^X = \left[\frac{\partial \log W_i^X}{\partial \log X_i} + 1\right]$ in expression B.6 that yields $\frac{\partial \log MC}{\partial \log \Omega} = -1 + \frac{\partial \log \phi(\boldsymbol{W})}{\partial \log \Omega} + \frac{\partial \log \mathcal{M}^X}{\partial \log \Omega}$. Thus, the presence of markdowns generates a positive bias in how marginal costs correlate with firm-size that is similar across all sectors. This is inconsistent with the negative correlation of marginal costs with firm-size documented in the homogeneous goods sector but positive association in the differentiated sector. Moreover, as described above, the strong positive correlation of markups using material inputs with those estimated using electricity suggests limited role of market power in the material input markets.

B.3 Appendix: Implications of Variable Markups for Reallocation Gains

This section complements Section 5 of main text. I first derive the relationship between the input demand X_i and any general firm subsidy S_i when markups are variable. I then derive a tax-subsidy policy S_i that guides the objective of a planner to reallocate resources while facing a fixed supply of aggregate resource.

B.3.1 Relation between input demand and tax-subsidy policy.

The change in input demand with respect to a change in subsidy is given by the total derivative of $\log X_i$:

$$d\log X_i \quad = \quad \left[\frac{\partial \log X_i}{\partial \log \mu_i}\frac{\partial \log \mu_i}{\partial \log S_i} + \frac{\partial \log X_i}{\partial \log \tau_i}\frac{\partial \log \tau_i}{\partial \log S_i}\right] d\log S_i \tag{B.15}$$

I use the fact that the subsidy (tax) reduces (increases) the marginal costs for a firm by exactly the amount of the tax-subsidy provided. That implies that $\frac{\partial \log \tau_i}{\partial \log S_i} = 1$. Next, I define firms' pass-through rate Γ_i as the elasticity of firm's price to its costs:

$$\Gamma_i \equiv \frac{\partial \log P_i}{\partial \log S_i} = \left[1 + \frac{\partial \log \mu_i}{\partial \log S_i}\right] \tag{B.16}$$

Substituting for these expression in B.15 yields:

$$d\log X_i = \left[\frac{\partial \log X_i}{\partial \log \mu_i}(\Gamma_i - 1) + \frac{\partial \log X_i}{\partial \log \tau_i}\right] d\log S_i \tag{B.17}$$

Because a input level can only be changed by varying the output level, the above expression can be rewritten as:

$$d\log X_i = \left[\underbrace{\frac{\partial \log X_i}{\partial \log Q_i}}_{=1/\theta_i^X}\underbrace{\frac{\partial \log Q_i}{\partial \log P_i}}_{=-\sigma_i}\left(\underbrace{\frac{\partial \log P_i}{\partial \log \mu_i}}_{=1}(\Gamma_i - 1) + \underbrace{\frac{\partial \log P_i}{\partial \log \tau_i}}_{=1}\right)\right] d\log S_i \tag{B.18}$$

where $\sigma_i = \mu_i/(\mu_i - 1)$ is the firms' demand elasticity. Using markup relationship from cost-minimization (equation 3) provides:

$$d\log X_i = -\left[\frac{\Gamma_i}{\theta_i^X - \alpha_i^X}\right] d\log S_i \tag{B.19}$$

B.3.2 Deriving a tax-subsidy policy.

I consider the marginal revenue product as the relevant margin of distortion. There are few advantages to do so. First, a large literature has now documented large dispersion in marginal product of inputs across manufacturing firms in developing countries (including India).[47] Viewed through a standard model of demand and supply, presence of distortions in marginal products is evidence of distortion. Second, it can be readily computed in the data using data on firms' revenue, input quantities, and estimates of output elasticity with respect to the inputs. The latter can be estimated using firm-level production data (as described in the B.2). Because distortions are not directly observed, marginal revenue product is considered a relevant summary statistic. Assuming the distortions are exogenous, a tax-subsidy policy can be devised using information on firms' marginal products. The policy should be such that the marginal product of input X is equalized across firms within an industry. Formally, let S_i represent the firm-level tax-subsidy (where $S_i > 1$ implies a tax and $S_i < 1$ implies as subsidy). Then S_i is defined such that:

$$S_i = \frac{\overline{\text{MRPX}}}{\text{MRPX}_i} \Rightarrow d\log S_i = d\log \text{MRPX}_i \tag{B.20}$$

where $d\log \text{MRPX}_i \equiv \log \overline{\text{MRPX}} - \log \text{MRPX}_i$. To obtain subsidy policy from the data requires knowledge of $\overline{\text{MRPX}}$. For this, I impose the constraint that the aggregate supply of resources X in the economy is fixed. This is the usual constraint imposed in the literature analyzing *static* misallocation. The constraint of fixed aggregate supply of resources implies that $\sum_i dX_i = 0$.

With the above objective and constraint, I can proceed with the calculation of tax-subsidy policy S_i. Using equation 4, the relation between (changes in) marginal revenue product and input demand is:[48]

$$d\log \text{MRPX}_i = (\theta_i^X - 1) d\log X_i$$

Define $\tilde{X}_i = \frac{X_i}{\theta_i^X - 1}$. This yields

$$dX_i = \tilde{X}_i \left(\log \overline{\text{MRPX}} - \log \text{MRPX}_i\right)$$

Summing over dX_i and using the aggregate supply of resources constraint $\sum_i dX_i = 0$ provides

[47] See Hsieh and Klenow (2009); Asker, Collard-Wexler, and De Loecker (2014); David and Venkateswaran (2019) and references therein.

[48] To see this, recall that $\text{MRPX} = P\frac{dQ}{dX} = P\theta^X \frac{Q}{X}$. Taking logs and first differences provides: $d\log \text{MRPX}_i = \left(\frac{d\log Q_i}{d\log X_i} - 1\right) d\log X_i$.

with the expression for the equalized $\overline{\text{MRPX}}$:

$$\log \overline{\text{MRPX}} = \left(\frac{\sum_i \tilde{X}_i \ . \ \log \text{MRPX}_i}{\sum_i \tilde{X}_i} \right) \tag{B.21}$$

Substituting for the expression of $\overline{\text{MRPX}}$ from B.21 back in equation B.20 gives us the tax-subsidy policy:

$$d \log S_i = \left[\left(\sum_i \left(\frac{\tilde{X}_i}{\sum_i \tilde{X}_i} \right) \ . \ \log \text{MRPX}_i \right) - \log \text{MRPX}_i \right]$$

C Empirical Appendix (for Online Publication)

C.1 A role for measurement error?

A potential concern with the analysis is that firm revenues or quantities might be measured with measurement error (ME). Notice that this is only an issue with the correlation results documented in section 3 and not for the estimates from identification strategy. In fact, an advantage of the identification strategy is that classical or non-classical ME will not affect the estimates because rain shocks are orthogonal to the error terms in estimated markup. I next address ME in correlations.

First, firms' employment might be positively associated with its prices if there are common reporting errors across the two variables across some years, generating positive bias in correlation estimates. To address this issue, I re-estimate the relation between firm's size and its costs and markups after instrumenting each firms' employment for every year using its initial employment (based on the first occurrence of each firm in the data) and its average employment across all years. The point estimates are virtually unaffected in these additional estimations. Second, I show that correlations documented in Figure I are robust to using the ranking of firms instead of using the levels. The use of ranks instead of levels relies less directly on the reported values and is less susceptible to correlations driven by outliers. Figure C.1 shows that the positive relations between costs and markups with firm size hold when using these ranking measures. Third, as documented in section 3, the correlation of marginal costs with firm size is of opposite signs across homogeneous and quality differentiated sector. Therefore, the correlation of ME across the two sectors will need to be of opposite sign to be able to explain positive relationship in differentiated sector and negative relationship of costs and size in homogeneous sector. This suggests that the correlations between firm size, markups and costs are not driven by ME bias, because such bias would have to also uniquely vary with quality differentiation.

C.2 Testing for alternative explanations for markup responses to demand shocks

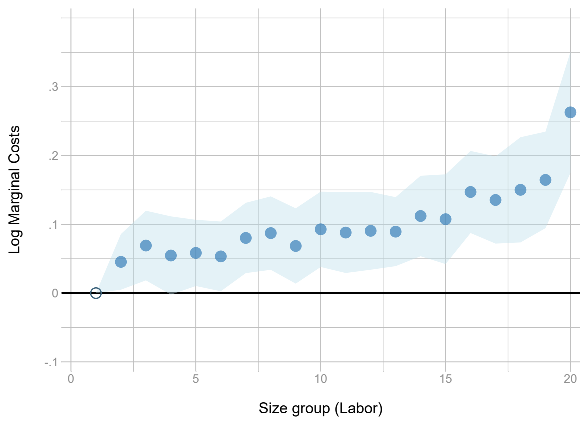
Firm entry and exit. Incumbent firms could lower their markup if new firms enter the market during high demand. This endogenous supply-side response to an increased demand increases the competition and exerts downward pressure on markups. I directly test for firm's entry and exit in the data. ASI data reports the year of establishment for firms as well as whether a firm is operational during the survey year. Table C.1 show that there is no evidence of excess entry or exit of firms in response to rain shocks.[49]

[49]Intuitively, firm entry or exit seems a remote candidate to drive the observed effects. Establishing a new firm requires substantial capital investment, labor hiring and it seems unlikely that firms would incur these large costs given the shifts in consumer demand induced by rain shocks are temporary in nature (Table E.7).

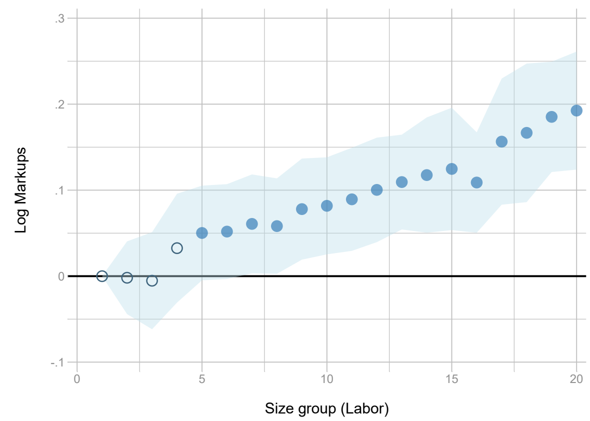
Figure C.1: Relation between firm's markups, marginal costs and size (by rankings)

Relationship by size groups

(a) Marginal costs

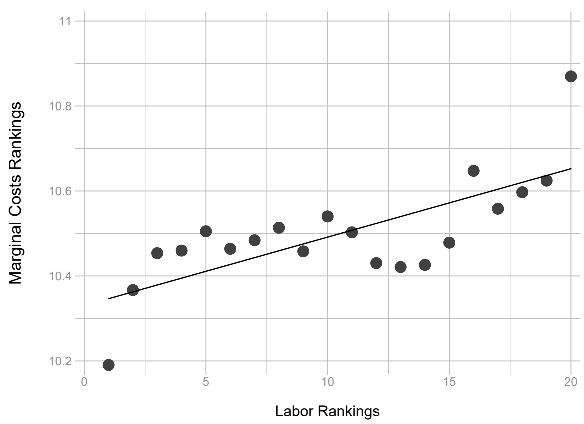


(b) Markups

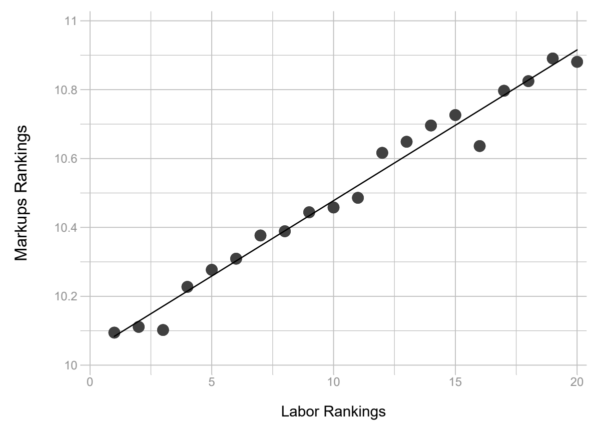


Relationship over discrete rankings

(c) Marginal costs



(d) Markups



The figure shows the relation between firm's per-unit marginal costs, markups and size by rankings across groups. Panel (a) and (b) show the relationship across firm size rankings, where a firms' ranking belongs to one of 20 size groups (based on the size of its labor force) within a district-product-year. Panel (c) and (d) also rank firms by their marginal costs and markups across 20 groups, and show how the rankings across marginal costs and markups relate to rankings across firm size.

New product introduction. Firms might introduce new products in response to higher demand, putting downward pressure on markups for the existing products Jaravel (2019). Two pieces of evidence suggest that this is unlikely the channel in my setting. First, it is the size of the market, and not the composition of the market, that matters for introduction of new products. Table C.2 rejects the hypothesis: effects of rain shocks on markups does not differ across districts with

Table C.1: Firm's entry/exit in response to rain shocks

	1(entry)	1(exit)	1(entry)	1(exit)
	(1)	(2)	(3)	(4)
Shock$_{dt}$(-1/0/+1)	0.001	0.000	-	-
	[0.001]	[0.001]	-	-
Shock$^{+}_{dt}$	-	-	-0.002	-0.001
	-	-	[0.002]	[0.001]
Shock$^{-}_{dt}$	-	-	-0.003	-0.002
	-	-	[0.002]	[0.001]
Observations	226,275	226,275	226,275	226,275
R-squared	0.358	0.312	0.358	0.312
Firm f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓

The table reports the estimates of new firm entry or incumbent exit based on specification: $1(entry/exit)_{it} = \alpha_i + \alpha_t + \beta$. Shock$_{dt} + \epsilon_{idt}$, where 1(entry) takes the value of 1 in the first year of firm's operation and 1(exit) takes the value of 1 when a firm is reported to be Closed in the survey. Shock$^{+}_{dt}$ and Shock$^{-}_{dt}$ takes the value of 1 if Shock$_{dt} = +1$ and Shock$_{dt} = -1$, and zero otherwise. Standard errors are clustered at the district level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

different *levels* of rural population. While the interaction with districts' *share* of rural population is significant, the estimate for interaction of rain shocks with district's *total* rural population is statistically insignificant. Second, the ASI data records product entry and exit, allowing me to test for this channel directly. Figure C.2 shows no effects on number of products across the firm-size distribution in response to rain shocks.

Collusion. In standard models of firm collusion, it is difficult to sustain collusion when demand changes frequently. This is because temptation to renege from a collusive agreement is higher during periods of temporary increase in demand because the gains from reneging are increasing in current demand but the loss from punishment increases in future (and uncertain) demand. If firms are indeed strategically adjusting their markups to build customer base then markups should decrease only in periods of higher demand. In periods with a drop in demand, however, markups should remain unchanged. The setting allows me to observe markup responses across both positive and negative demand shocks. Figure C.3 confirms that the non-monotonic effects of rain shocks on markups are present across both positive and negative rain shocks. Therefore, the prediction from models of firm collusion does not hold support in the data.

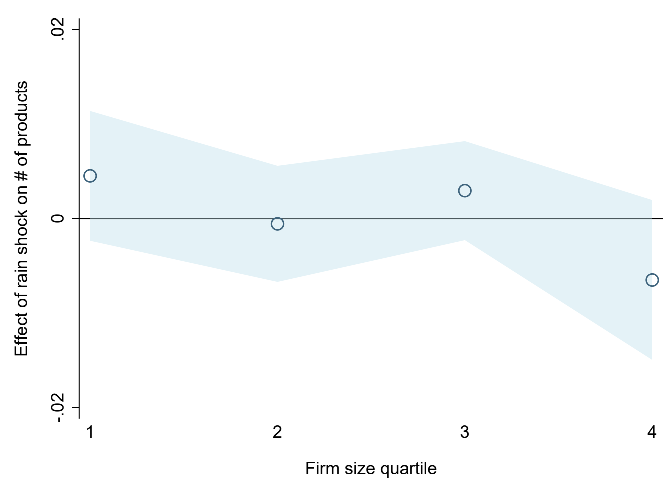
Consumer search. Consumers might increase their search intensity and shop more outlets during periods of high demand, appearing to be more price sensitive to firms. While both increased consumer search and changes in demand composition would affect markups, they emphasize

Table C.2: Composition effect versus size effect

	Dependent variable: log (markup)		
	(1)	(2)	(3)
Shock$_{dt}$ (-1/0/+1)	-0.002	-0.010*	-0.010*
	[0.002]	[0.005]	[0.005]
Shock$_{dt}$ $\times$ **1**(High Share of rural pop.)$_d$	-0.006**	-	-0.007**
	[0.003]	-	[0.003]
Shock$_{dt}$ $\times$ log(Total rural population)$_d$	-	0.000	0.001
	-	[0.000]	[0.000]
Observations	133,094	133,094	133,094
R-squared	0.989	0.989	0.989
Firm-product f.e.	✓	✓	✓
Product-year f.e.	✓	✓	✓

The table reports the effects of rain shocks on markups, by share of rural population and total rural population. 1(High share of agricultural population)$_d$ takes the value of 1 if more than two-third of districts' population is rural as reported in the 2001 Census of India. Total rural population for the district is sourced from the 2001 Census of India. All columns include firm-product, product-year fixed effects and control for log marginal costs. Standard errors clustered at district level are reported in brackets. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

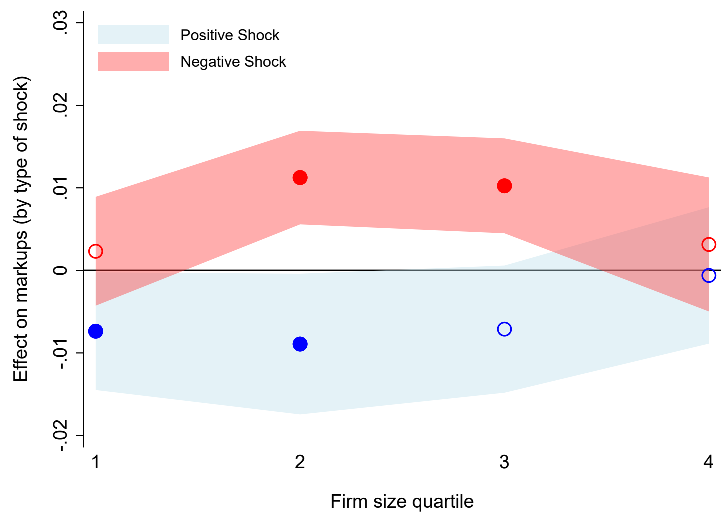
Figure C.2: Effect of rain shocks on number of products



The figure reports the heterogeneous effects of rain shocks on number of products based on specification: $\log y_{it} = \sum_{r=1}^{4} \beta^r.\ (\text{Shock}_{dt} \times Q_i^r) + \alpha_i + \alpha_{kt} + \epsilon_{it}$, where y_{it} are the number of products for firm i in year t. Shock$_{dt}$ is as defined in the main text.

different mechanisms due to which firms would lower markups when demand increases. Under the consumer search channel, time-varying demand elasticity faced by firms is a result of increased search activity. As a result, higher search intensity in periods of increased demand would predict a positive association between a firm's demand elasticity and changes to its demand. In the context

Figure C.3: Effect on markups by positive and negative shocks



The figure reports the heterogeneous effects of positive and negative rain shocks on markups from the specification: $\log \mu_{ijt} = \sum_{r=1}^{4} \beta^{r,+}.\ (\text{Shock}_{dt}^{+} \times Q_i^r) + \sum_{r=1}^{4} \beta^{r,-}.\ (\text{Shock}_{dt}^{-} \times Q_i^r) + \alpha_{ij} + \alpha_{jt} + \Gamma' X_{ijt} + \epsilon_{ijt}$, where Shock_{dt}^{+} and Shock_{dt}^{-} takes the value of 1 if $\text{Shock}_{dt} = +1$ and $\text{Shock}_{dt} = -1$, and zero otherwise. Specification includes firm-product, product-year fixed effects and controls for firm age, size quartile-year fixed effects, and log marginal costs. 95% confidence intervals are represented by shaded area. Bold circles indicate estimates significant at the 10% level, and hollow circles statistically insignificant from 0 at the 10% level.

of this paper, this implies that smallest firms should see the largest increase in their price elasticity and lower their markups. However, I find that markup responses are only present for firms in the middle of the size distribution.

Financial constraints. Firms facing costly external financing may raise their markups when faced with negative demand shocks Gilchrist, Schoenle, Sim, and Zakrajsek (2017). In these models, consumers have persistent habit over firms' products. This allows financially distressed firms to increase their markups and increase cash holdings, allowing them to avoid liquidation in the short-run. Two results rule out the financial channel as a potential driver for the results. First, as reported in Table C.3, the estimates of interaction of rain shocks with firm-size are robust when I include as controls the differential effect of rain shocks depending on firm's financial strength, proxied by firm's cash ratio $\left(\frac{\text{Cash}}{\text{Cash + Fixed Assets}}\right)$, and it's financial leverage $\left(\frac{\text{Debt}}{\text{Fixed Assets}}\right)$. Second, as documented previously, markups for smallest firms do not change in response to negative rain shocks. Smaller firms are more likely to have binding financial constraints, and therefore, the results on no effect on markups for these firms is in contrast with a financial constraint channel.

Table C.3: Robustness to financial frictions

	Dependent variable: log (markup)		
	(1)	(2)	(3)
Shock$_{dt}$ (-1/0/+1)			
× First size quartile	-0.003	-0.004	-0.003
	[0.003]	[0.003]	[0.003]
× Second size quartile	-0.009***	-0.009***	-0.009***
	[0.003]	[0.003]	[0.003]
× Third size quartile	-0.007**	-0.007**	-0.007**
	[0.003]	[0.003]	[0.003]
× Fourth size quartile	0.001	0.001	0.001
	[0.004]	[0.004]	[0.004]
Shock$_{dt}$ (-1/0/+1)			
× Cash Ratio	-0.003		-0.003
	[0.015]		[0.015]
× Leverage		-0.000	-0.000
		[0.000]	[0.000]
Observations	132,746	132,746	132,746
R-squared	0.989	0.989	0.989
Firm-product f.e.	✓	✓	✓
Product-year f.e.	✓	✓	✓

The table tests for robustness of estimates after controlling for differential effects of rain shocks on firms' financial strength. All specifications include firm-product, product-year fixed effects and control for firm age, size quartile-year fixed effects and for log marginal costs. Coefficients on levels of financial strength are not reported for brevity. Standard errors clustered at district level are reported in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

D Comparison with existing models on firm heterogeneity (For Online Publication)

In this section, I compare the cross-sectional and time-series predictions from existing models from literature that feature firm heterogeneity and variable markups. For cross-sectional predictions, I compare the relationship between firm-size, with their markups and marginal costs. For time-series predictions, I compare the predictions on how firms across the size distribution would change their markups in response to demand shocks to the poor households. These predictions are summarized in Table D.1.

D.1 Models with monopolistic competition

Melitz (2003) is the benchmark efficiency sorting model with CES demand and this framework more efficient firms have lower marginal costs. A number of studies have incorporated Melitz (2003) framework with quality differentiation. Under these models, more productive and larger firms charge higher prices for their products. This higher price is a premium for quality and is driven by higher marginal costs: production of better quality entails expensive and better quality inputs (Kugler and Verhoogen 2011).

Under CES demand, however, all firms optimally charge a constant markup over marginal costs. Therefore, markups do not vary with firm size, and neither do they vary across time in response to demand shocks in either efficiency sorting or quality sorting frameworks.

Melitz and Ottaviano (2008) present an efficiency sorting model where firms face linear demand. Unlike CES preferences, the price elasticity of demand faced by firms is not constant in these models but rather depends on degree of competition among firms in these markets. Firms facing lower competition charge higher markups. Efficiency sorting implies that larger firms have lower marginal costs and offer lower prices, even though they have higher markups.

Zhelobodko, Kokovin, Parenti, and Thisse (2012) propose a variable markups framework with endogenous consumer demand elasticity. In their model, consumers have higher preference over larger varieties and consume more variety as their income increases. An increase in variety increases their demand elasticity and lowers firms' markups. Under this framework, we should expect the markups to decrease the most for smallest firms in response to an increase in demand from the poor.

Kneller and Yu (2016) embed quality differentiation in Melitz and Ottaviano (2008) framework. Firms with higher costs produce better quality and charge higher markup as they are able to command larger market share. However, in response to an increase in demand — irrespective of the income group from which demand increases — the markups increase.

D.2 Models with imperfect competition

In framework of Atkeson and Burstein (2008) and Edmond, Midrigan, and Xu (2015), firms face CES demand and compete in a olipolistic competitive market. Larger firms are more efficient and have lower costs. Larger firms also charge higher markups as they command higher market shares. An increase in demand from poor households, however, increases the market share for smallest and mid-size firms and they should increase their markups. Market share for largest firms shrink and they should lower their markups. Bastos, Silva, and Verhoogen (2018) incorporate quality sorting in Atkeson and Burstein (2008) framework. Due to quality sorting, costs increase with firm size in Bastos, Silva, and Verhoogen (2018). Relation between markups and firm size is similar to that in Atkeson and Burstein (2008).

Table D.1: Existing models of firm heterogeneity and variable markups

		Correlation between firm size and		Effect of Δ_t Demand from the poor on Δ_t Markup for ..		
Nature of firm heterogeneity, competition, and demand	Relevant Papers	Marginal Cost (1)	Markups (2)	Smallest firms (3)	Mid-size firms (4)	Largest firms (5)
Efficiency sorting, Monopolistic Comp., and CES	Melitz (2003)	-	0	0	0	0
Quality sorting, Monopolistic Comp., and CES	Verhoogen (2008) Kugler and Verhoogen (2011) Hallak and Sivadasan (2013)	+	0	0	0	0
Efficiency sorting, Monopolistic Comp., and non-CES	Melitz and Ottaviano (2008) Zhelobodko et. al. (2012) Edmond, Midrigan, and Xu (2019)	-	+	- - -	- -	0
Quality sorting, Monopolistic Comp., and Linear	Kneller and Yu (2016)	+	+	+++	++	0
Efficiency sorting, Oligopolistic Comp., CES	Atkeson and Burstein (2008) Edmond, Midrigan, and Xu (2015)	-	+	+++	+/-	- - -
Quality sorting, Oligopolistic Comp., CES	Bastos, Silva, and Verhoogen (2018)	+	+	+++	++	-
Quality sorting, heterogeneous demand elasticities	This paper and data	+	+	-	- - -	-

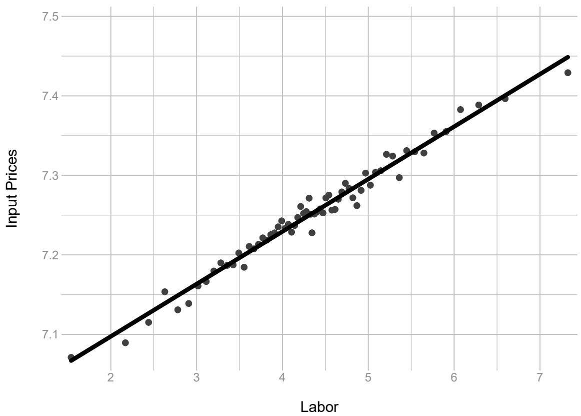
The number of signs reflect the relative intensity of effects in Column 3-5. For example, +++ implies that the positive effect is higher when compared to ++, which is higher than +.

Taken together, this exercise shows that existing models cannot explain the three findings of the paper in combination: (a) larger firms have higher marginal costs (b) larger firms have higher markups (c) markups are decreasing for mid-sized firms in response to an increase in demand from the poor.

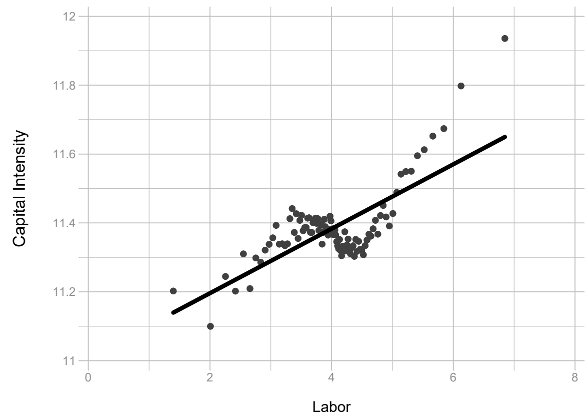
E Appendix Figures and Tables (For Online Publication)

Figure E.1: Relation between firm size and input factor costs

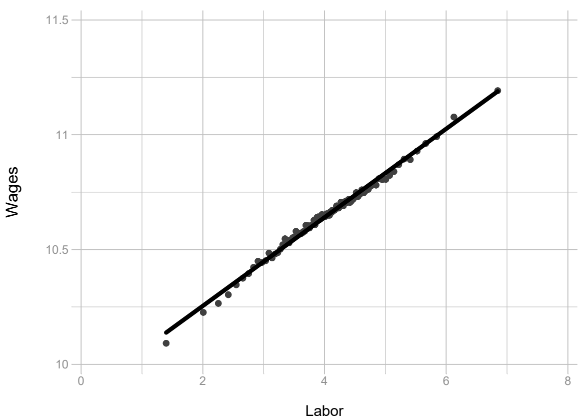
(a) log input prices



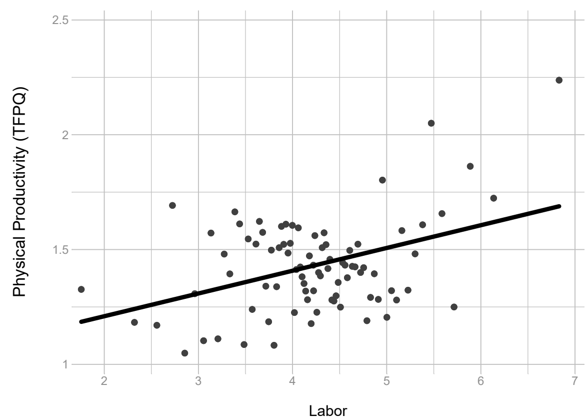
(b) log Capital/Labor



(c) log wages

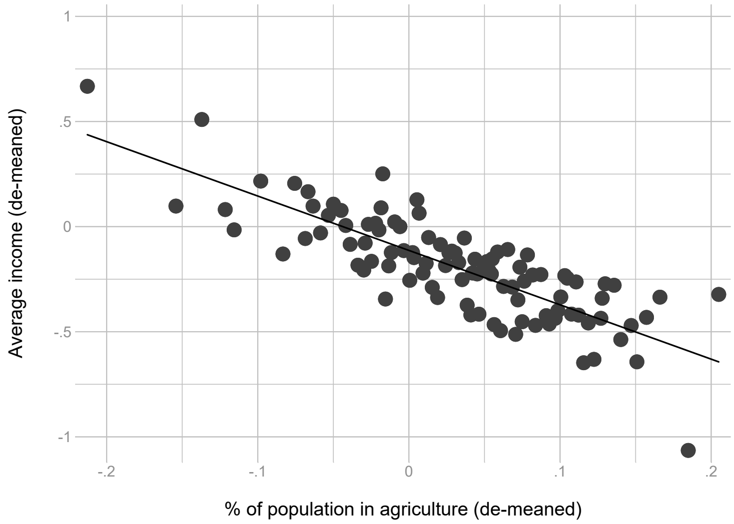


(d) log physical productivity



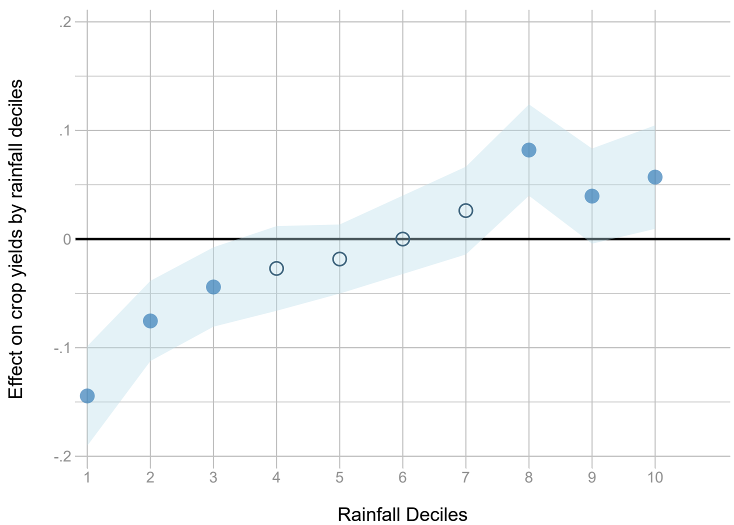
The figure shows the relation between firm size (as measured by its labor force) and input prices (Panel (a)), capital intensity (Panel (b)), wages per unit labor (Panel (c)), physical productivity TFPQ (Panel (d)). All variables are measured in logs. All specifications control for district-by-product-by-year fixed effects. Each dot represents 1% of observations. Source: ASI

Figure E.2: Share of agricultural population and average income in district



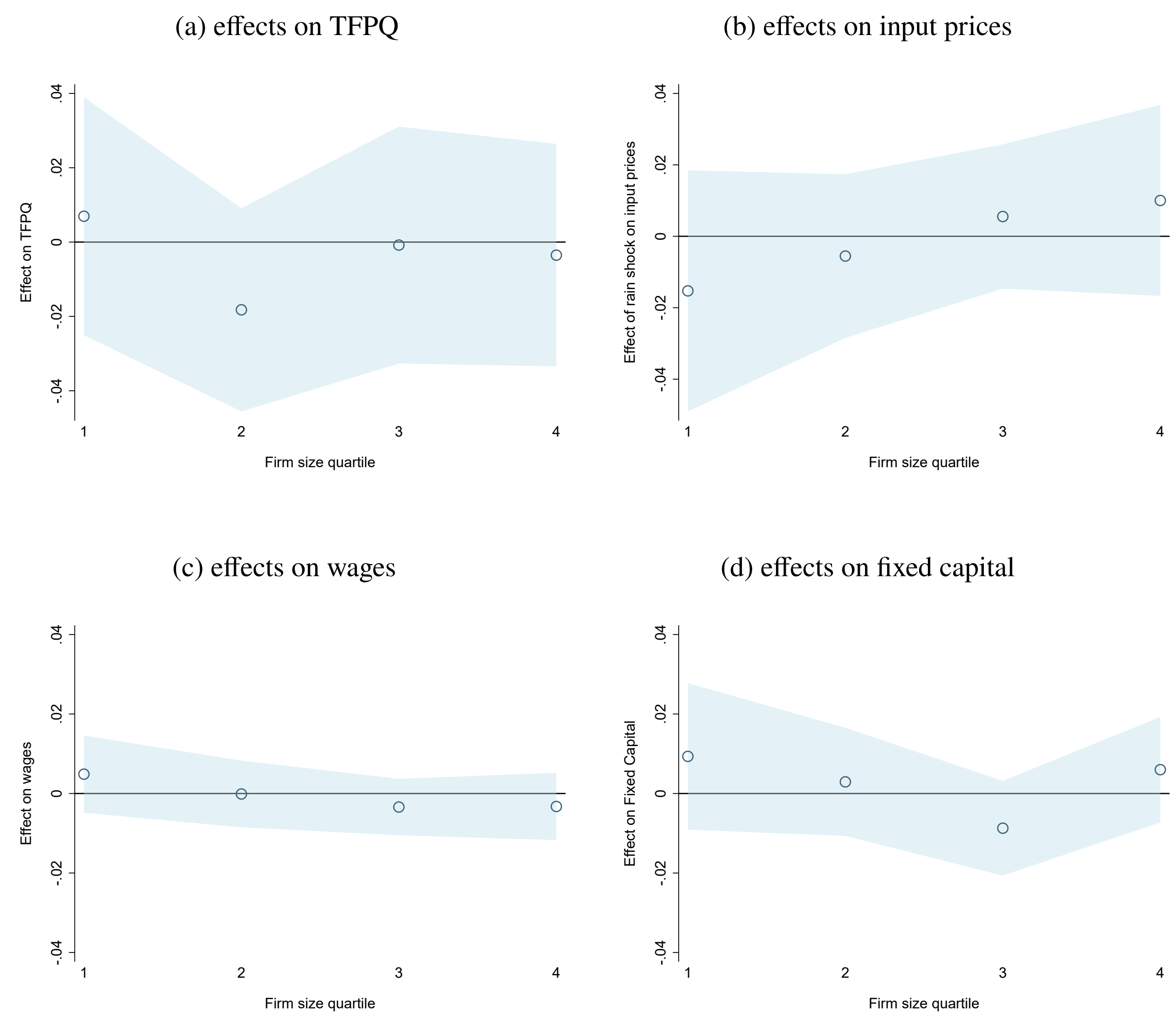
The figure plots the relation between share of population involved in agricultural activities and average income in the district. Both axes plot the residualized values after removing state fixed effects. The correlation is -4.52 and is significant at 1% levels (t = -5.39) when standard errors clustered at district level. Source: NSS

Figure E.3: Effect of rain shocks on agricultural yields



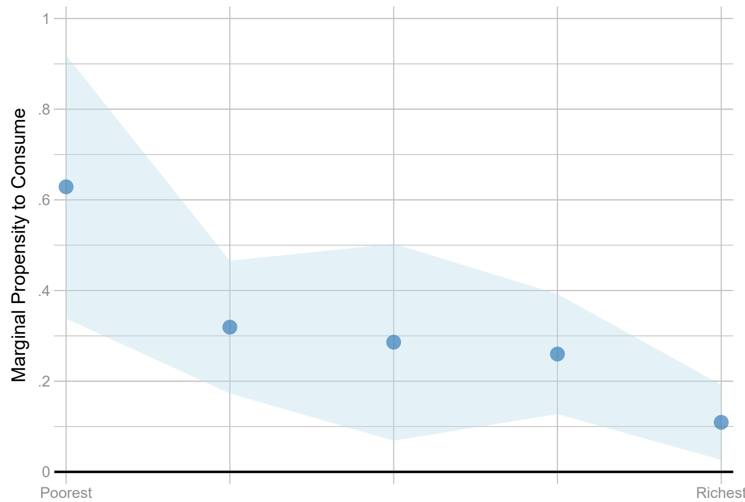
The figure plots coefficients and 95% confidence intervals from a regression of log crop yields on dummies for each decile of the rainfall distribution within the district. Log crop yields is the log of a weighted average of yields of the 15 crops for which data is available in the VDSA database. The yield for each crop has first been normalized by the mean yield of that crop in the district. Weights are the mean percentage of land area planted with a given crop in a district. Each decile dummy equals 1 if monsoon rainfall in the current year fell within the given decile of the district's usual rainfall distribution for that year and equals 0 otherwise. The omitted category against is the 6^{th} decile. Regression specification includes district and year fixed effects. Standard errors are clustered at district level.

Figure E.4: Effect of rain shocks on TFPQ, input prices, wages and capital



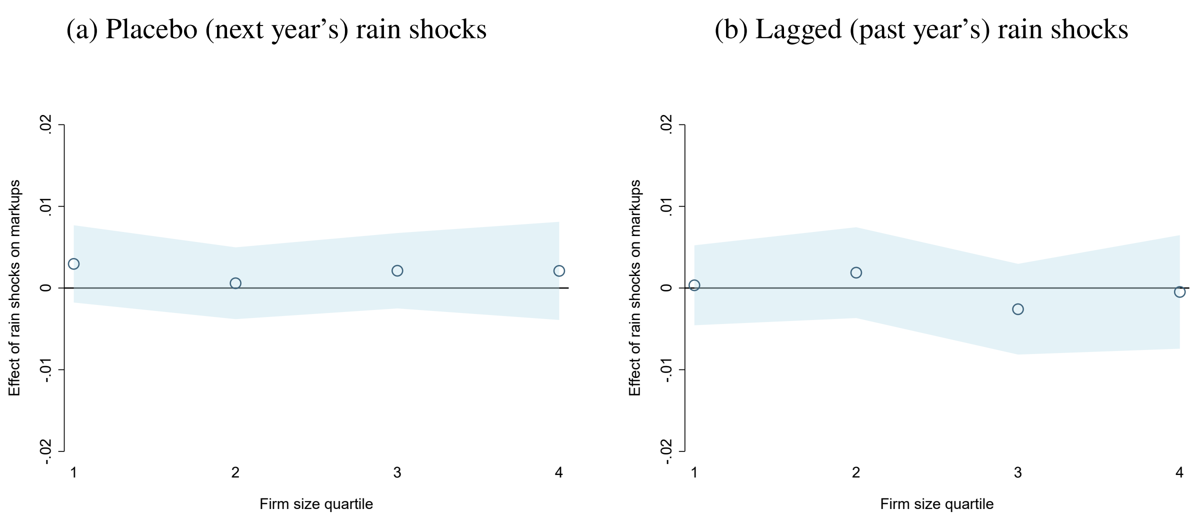
The figure shows the estimates of the effect of rain shocks across the firm-size distribution on TFPQ (Panel (a)), input prices (Panel (b)), wages per unit labor (Panel (c)), fixed capital (Panel (d)). All dependent variables are measured in logs. All specifications control for firm age and size quartile-year fixed effects. 95% confidence intervals are represented by shaded blue area. Bold circles indicate results that are significant at the 10 level, and hollow circles statistically insignificant from 0 at the 10% level.

Figure E.5: Marginal Propensity to Consume (MPC) across income groups



The figure reports the estimate of marginal propensity to consume (MPC) across income groups. It plots the estimates $\alpha(z)$ across five income groups based on the following specification: $\Delta \log x_{ivt}(z) = \alpha(z) \Delta \log y_{ivt}(z) + \beta_i + \gamma_{vt} + \epsilon_{ivt}$ where β_i is the household fixed effect and γ_{vt} is a town-year fixed effects. Changes in employment status are used as an instrument for changes in income. Source: CMIE

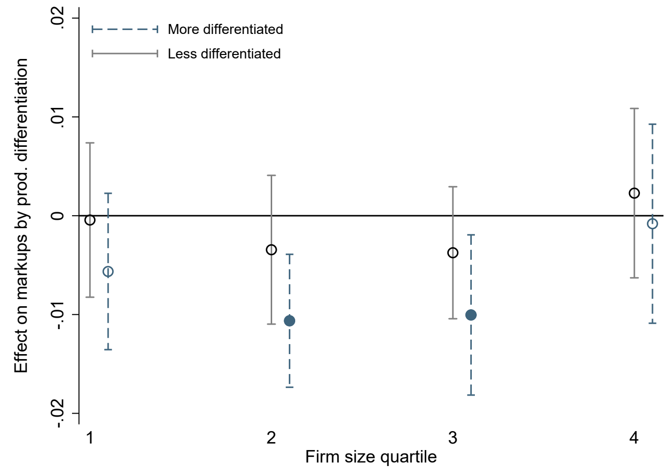
Figure E.6: Effects of placebo and past rain shocks on markups



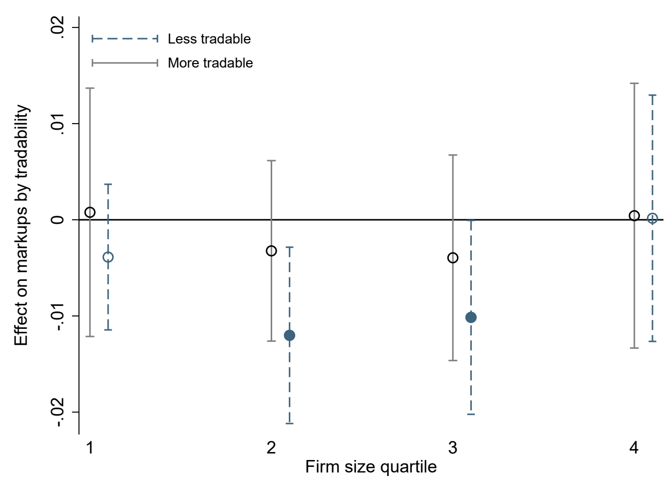
95% confidence intervals are represented by shaded blue area. Bold circles indicate results that are significant at the 10 level, and hollow circles statistically insignificant from 0 at the 10% level.

Figure E.7: Effect of rain shocks on markups across firm-size distribution (by product differentiation and industry tradability)

(a) by quality differentiation



(b) by industry tradability



The figure shows the estimates of the effect of rain shocks on markups across the firm-size distribution by scope of quality-differentiation (Panel (a)) and tradability (Panel (b)) using specification 12. All specification includes firm-product, product-year fixed effects and controls for firm age, size quartile-year fixed effects, and log marginal costs. 95% confidence intervals are represented by vertical lines in both panels. Bold circles indicate estimates significant at the 10% level, and hollow circles statistically insignificant from 0 at the 10% level.

Table E.1: Summary Statistics: Distribution of sales across industries

	Share of Output		Share of Output
15 Food and beverages	23%	26 Non-metal minerals	8%
16 Tobacco products	2%	27 Basic Metals	15%
17 Textiles	8%	28 Fabricated metal	1%
18 Wearing apparel	1%	29 Machinery	5%
19 Leather products	1%	31 Electric	2%
20 Wood products	0%	32 Communications prod.	1%
21 Paper products	1%	33 Medical equipment	0%
22 Printing	0%	34 Motor vehicles	6%
23 Coke products	7%	35 Other transport	4%
24 Chemicals	11%	36 Furniture	1%
25 Rubber and Plastic	3%		

The table reports the share of total output by 2-digit industries (averaged across years) in the Annual Survey of Industries (ASI) data.

Table E.2: Baseline correlations using alternative measures of firm size

	Dependent variable: log of ...				
Panel A:	Marg. Cost	Markup	Material Inputs	K/L	Wages
	(1)	(2)	(3)	(4)	(5)
(log) sales	0.020**	0.055***	0.048***	0.246***	0.164***
	[0.008]	[0.009]	[0.011]	[0.009]	[0.006]
R-squared	0.870	0.639	0.410	0.682	0.830
Panel B.	(1)	(2)	(3)	(4)	(5)
(log) assets	-0.011	0.067***	0.049***	0.614***	0.119***
	[0.007]	[0.004]	[0.008]	[0.015]	[0.005]
R-squared	0.870	0.641	0.410	0.879	0.809
Observations	167,221	167,221	443,022	167,221	167,221
Industry f.e.	✓	✓	✓	✓	✓
District-prod.-year f.e.	✓	✓	✓	✓	✓

The table reports the correlation from Table I using alternate definition of firm size based on total sales (Panel A) and total fixed assets (Panel B). Standard errors clustered by district level are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table E.3: Estimates of price-elasticity of demand (σ)

	OLS	IV	IV
	(1)	(2)	(3)
(1- σ) All households	-0.095***	-0.106**	-
	[0.027]	[0.041]	-
(1- σ) Poorest Quintile (Relative to Richest)	-	-	-0.161***
			[0.022]
(1- σ) 2nd poorest Quintile (Relative to Richest)	-	-	-0.123***
			[0.017]
(1- σ) Median Quintile (Relative to Richest)	-	-	-0.089***
			[0.013]
(1- σ) 2nd richest Quintile (Relative to Richest)	-	-	-0.050***
			[0.008]
F-stat		223.601	45.035
Observations	423,864	423,864	423,864
Region f.e.	✓	✓	✓
Quintile f.e.	✓	✓	✓

The table reports the estimate of price-elasticity of demand based on the estimating equation 5. Column 2-3 estimates are based on the IV specification that instruments change in price of a good with state-level leave out mean price changes (described in Section 4.2). Standard errors clustered at district level are reported in brackets. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table E.4: Estimates of price-elasticity of demand across product groups

	Estimates of (1- σ) across the product category:						
	Vegetables (1)	Fruits (2)	Dry Fruits (3)	Spices (4)	Tobacco (5)	Footwear (7)	Clothes (8)
Poorest Quintile (Relative to Richest)	-0.025**	-0.052***	-0.075***	-0.035*	-0.107***	-0.011	-0.083***
	[0.012]	[0.012]	[0.027]	[0.019]	[0.020]	[0.009]	[0.018]
2nd poorest Quintile (Relative to Richest)	-0.021*	-0.058***	-0.086***	-0.034**	-0.123***	-0.015**	-0.077***
	[0.011]	[0.012]	[0.019]	[0.015]	[0.021]	[0.007]	[0.015]
Median Quintile (Relative to Richest)	-0.010	-0.043***	-0.065***	-0.016	-0.095***	-0.010*	-0.056***
	[0.008]	[0.010]	[0.016]	[0.011]	[0.017]	[0.006]	[0.011]
2nd richest Quintile (Relative to Richest)	-0.003	-0.035***	-0.044***	-0.012*	-0.098***	-0.006	-0.037***
	[0.006]	[0.009]	[0.011]	[0.007]	[0.016]	[0.005]	[0.007]
Observations	909,149	113,360	51,409	321,830	56,609	705,341	155,459
Region-product f.e.	✓	✓	✓	✓	✓	✓	✓
Quintile f.e.	✓	✓	✓	✓	✓	✓	✓

The table reports the estimate of price-elasticity of demand based on the estimating equation 5 across multiple product groups. All estimates are based on the IV specification that instruments change in price of a good with state-level leave out mean price changes (described in Section 4.2). Standard errors clustered at district level are reported in brackets. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table E.5: Rainfall induced income shocks for poor population

	Dependent variable: log of ...				
	Agricultural output		Daily wages		
	Crop yield (1)	Revenue per unit area (2)	Rural agri. labor (3)	Rural non-. -agri labor (4)	Non-rural labor (5)
Shock$_{dt}$ (-1/0/+1)	0.045***	0.035***	0.027***	-0.009	0.017
	(0.005)	(0.005)	[0.008]	[0.009]	[0.011]
Observations	38,280	38,280	115,852	102,910	154,939
R-squared	0.887	0.853	0.516	0.271	0.124
District-crop f.e.	✓	✓			
Crop-year f.e.	✓	✓			
District f.e.			✓	✓	✓
Year f.e.			✓	✓	✓

The table reports the effect of rain shocks on agricultural productivity and labor market. Standard errors clustered by district level are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table E.6: Estimates of price elasticities across industries

	OLS	IV		OLS	IV
Sector	(1)	(2)	Sector	(1)	(2)
Pooled (Average)	-0.668***	-2.364***	Rubber and plastic	-0.634***	-2.916***
	[0.011]	[0.044]		[0.025]	[0.124]
Food and beverages	-0.472***	-4.543***	Non-metal minerals	-0.580***	-2.305***
	[0.022]	[0.168]		[0.024]	[0.080]
Tobacco products	-0.281***	-2.617***	Basic Metals	-0.641***	-2.060***
	[0.104]	[0.317]		[0.015]	[0.051]
Textiles	-0.530***	-4.462***	Fabricated metal	-0.675***	-2.014***
	[0.028]	[0.174]		[0.015]	[0.049]
Wearing apparel	-0.465***	-3.900***	Machinery	-0.783***	-2.011***
	[0.056]	[0.194]		[0.013]	[0.045]
Leather products	-0.510***	-3.438***	Electric	-0.747***	-1.919***
	[0.063]	[0.257]		[0.016]	[0.045]
Wood products	-0.726***	-2.348***	Communications	-0.733***	-2.033***
	[0.035]	[0.088]		[0.026]	[0.062]
Paper products	-0.718***	-2.559***	Medical equipment	-0.820***	-2.104***
	[0.032]	[0.136]		[0.026]	[0.061]
Printing	-0.709***	-2.534***	Motor vehicles	-0.626***	-1.986***
	[0.035]	[0.119]		[0.019]	[0.051]
Coke products	-0.460***	-3.241***	Other transport	-0.643***	-1.993***
	[0.039]	[0.136]		[0.018]	[0.050]
Chemical	-0.517***	-3.520***	Furniture	-0.702***	-2.531***
	[0.019]	[0.120]		[0.026]	[0.077]

The table reports the estimated of price elasticities from specification 9. Columns 2 report IV estimates where price is instrumented with marginal costs. All specifications include product-year fixed effects. Standard errors are clustered by firm-product level (N = 133,094). Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table E.7: Testing for serial correlation in rainfall

	Dependent variable: RainDeviation$_{d,t}$			
	1998-2009 (Sample Years)		1990-2014	
	(1)	(2)	(3)	(4)
RainDeviation$_{d,t-1}$	-0.007	-0.007	-0.013	-0.014
	(0.029)	(0.028)	(0.015)	(0.015)
RainDeviation$_{d,t-2}$	-	0.007	-	0.016
	-	(0.030)	-	(0.012)
Observations	3,116	3,116	7,850	7,850
R-squared	0.231	0.231	0.261	0.261
District f.e.	✓	✓	✓	✓
Year f.e.	✓	✓	✓	✓

This table tests for serial correlation in rainfall. The estimates are based on the following specification: RainDeviation$_{dt}$ $= \alpha_d + \alpha_t + \beta_1$ RainDeviation$_{d,t-1}$ $+ \beta_2$ RainDeviation$_{d,t-2}$ $+ \epsilon_{dt}$, where RainDeviation$_{dt}$ is the rainfall deviation in district d and year t form the median rainfall of the district since 1960. Standard errors are clustered at the district level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table E.8: Robustness to definition of rain shocks

	Percentile cut-off for Positive/Negative Shocks				Deviations from
	80/20	80/30	85/15	90/10	the median
	(1)	(2)	(3)	(4)	(5)
	Panel A. Dependent variable: log (markup)				
Shock$_{dt}$	-0.005**	-0.005***	-0.005**	-0.005**	-0.002***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]
R-squared	0.989	0.989	0.989	0.989	0.989
	Panel B. Dependent variable: log (marginal costs)				
Shock$_{dt}$	0.007	0.002	0.008	0.007	0.003
	[0.007]	[0.007]	[0.008]	[0.010]	[0.003]
R-squared	0.952	0.952	0.952	0.952	0.952

The table shows the estimates from specification 8 based on alternate definitions of rain shocks. In Column 1-4 I use different cut-offs of rain shocks in equation 8: positive and negative shocks are defined as rain shocks above/below 80/20, 80/30, 85/15 and 90/10 percentiles. In Column 5, I use continuous measure of rain shock defined as rainfall deviation relative to the historical rainfall received in the district. All specifications include firm-product and product-year fixed effects. Standard errors clustered at district level are reported in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1. (N = 133,094)

Table E.9: Effect of rain shocks on exporters

	Dependent variable: log of ...					
	quantity	markup	marg. cost	quantity	markup	marg. cost
	(1)	(2)	(3)	(4)	(5)	(6)
Shock$_{dt}$ (-1/0/+1)	0.015	0.019	-0.010	-	-	-
	[0.023]	[0.027]	[0.030]	-	-	-
Rain deviations	-	-	-	0.012	0.006	-0.002
from median$_{dt}$	-	-	-	[0.010]	[0.010]	[0.012]
Firm-product f.e.	✓	✓	✓	✓	✓	✓
Product-year f.e.	✓	✓	✓	✓	✓	✓

Notes: The table analyzes the effect of rain shocks on exporters. Standard errors clustered at district level are reported in parenthesis. Significance: *** p<0.01, ** p<0.05, * p<0.1. (N=10,114)

Table E.10: Effects of rain shocks on markups across firm-size distribution

	Dependent variable: log markup							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Shock$_{dt}$ (-1/0/+1)								
× First size quartile	-0.003	-0.003	-0.003	-0.004	-0.004	-0.002	-0.001	-0.002
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.003]
× Second size quartile	-0.009***	-0.010***	-0.009***	-0.009***	-0.010***	-0.008**	-0.007**	-0.008**
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]
× Third size quartile	-0.007**	-0.005*	-0.007**	-0.007**	-0.008***	-0.006*	-0.005	-0.006*
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
× Fourth size quartile	0.001	0.000	0.001	0.001	0.001	0.003	0.003	0.003
	[0.003]	[0.004]	[0.003]	[0.004]	[0.003]	[0.004]	[0.004]	[0.004]
Observations	133,094	122,828	133,094	133,094	133,094	133,094	133,094	133,094
R-squared	0.989	0.990	0.989	0.989	0.989	0.989	0.989	0.989
Firm-product f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Product-year f.e.	✓	✓	✓	✓	✓	✓	✓	✓
Controls	Baseline Specification	Single-plant firms	+ Age control	+ Size-year control	Past 2-year shocks controls	National Market access control	In + out-state market access	(3)-(7) controls

The table reports effects of rain shocks on markups across the firm-size distribution (β^r from specification 10). Shock$_{dt}$ is as defined in the text. All columns include firm-product, product-year fixed effects and control for log marginal costs. Standard errors clustered by district level are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table E.11: Estimates of pass-through rates

	Dependent variable: $\log \text{price}_{ijt}$					
	OLS estimation			IV estimation		
	(1)	(2)	(3)	(4)	(5)	(6)
$\log \text{mc}_{ijt}$	0.553***	0.678***	0.664***	0.700***	0.739***	0.855***
	[0.006]	[0.015]	[0.025]	[0.008]	[0.023]	[0.036]
$\log \text{mc}_{ijt} \times \log \text{labor}_{it}$		-0.026***	-0.019***		-0.009*	-0.014*
		[0.003]	[0.005]		[0.005]	[0.008]
$\log \text{mc}_{ijt} \times 1(\text{diff})_i$			0.030			-0.166***
			[0.031]			[0.046]
$\log \text{mc}_{ijt} \times \log \text{labor}_{it} \times 1(\text{diff})_i$			-0.014**			0.004
			[0.007]			[0.010]
Observations	131,557	131,557	131,557	131,557	131,557	131,557
R-squared	0.408	0.411	0.411			
Kleibergen-Paap F-stat				8392.57	847.44	132.57
Firm-product f.e.	✓	✓	✓	✓	✓	✓
NIC4 - year f.e.	✓	✓	✓	✓	✓	✓

The table reports estimates of pass-through rates from specification 19. Standard errors clustered by firm-level are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Appendix References (For Online Publication)

Asker, J., A. Collard-Wexler, and J. De Loecker (2014): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122(5), 1013–1063.

Atkeson, A., and A. Burstein (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98(5), 1998–2031.

Atkin, D., A. K. Khandelwal, and A. Osman (2017): "Exporting and Firm Performance: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 132(2), 551–615.

Bastos, P., J. Silva, and E. Verhoogen (2018): "Export Destinations and Input Prices," *American Economic Review*, 108(2), 353–92.

Comin, D., D. Lashkari, and M. Mestieri (2021): "Structural Change With Long-Run Income and Price Effects," *Econometrica*, 89(1), 311–374.

David, J. M., and V. Venkateswaran (2019): "The Sources of Capital Misallocation," *American Economic Review*, 109(7), 2531–67.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, Markups, and Trade Reform," *Econometrica*, 84(2), 445–510.

De Loecker, J., and F. Warzynski (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102(6), 2437–71.

Edmond, C., V. Midrigan, and D. Y. Xu (2015): "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105(10), 3183–3221.

——— (2019): "How Costly Are Markups?," *Working Paper*.

Faber, B., and T. Fally (2020): "Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data," *Working Paper*.

Garcia-Marin, A., and N. Voigtländer (2019): "Exporting and Plant-Level Efficiency Gains: It's in the Measure," *Journal of Political Economy*, 127(4), 1777–1825.

Gilchrist, S., R. Schoenle, J. Sim, and E. Zakrajsek (2017): "Inflation Dynamics during the Financial Crisis," *American Economic Review*, 107(3), 785–823.

Hallak, J., and J. Sivadasan (2013): "Product and process productivity: Implications for quality choice and conditional exporter premia," *Journal of International Economics*, 91(1), 53–67.

Handbury, J. (2019): " Are poor cities cheap for everyone? Non-homotheticity and the cost of living across US cities," *Working Paper*.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Jaravel, X. (2019): "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *The Quarterly Journal of Economics*.

Kim, R. (2016): " Price-Cost Markup Cyclicality: New Evidence and Implications," *Working Paper*.

Kneller, R., and Z. Yu (2016): "Quality Selection, Sectoral Heterogeneity and Chinese Exports," *Review of International Economics*, 24(4), 857–874.

Kugler, M., and E. Verhoogen (2011): "Prices, Plant Size, and Product Quality," *The Review of Economic Studies*, 79(1), 307–339.

Mayer, T., M. J. Melitz, and G. I. P. Ottaviano (2014): "Market Size, Competition, and the Product Mix of Exporters," *American Economic Review*, 104(2), 495–536.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Melitz, M. J., and G. I. P. Ottaviano (2008): "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 75(1), 295–316.

Verhoogen, E. A. (2008): "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 123(2), 489–530.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012): "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," *Econometrica*, 80(6), 2765–2784.