Competition, Financial Constraints, and Misallocation: Plant-Level Evidence from Indian Manufacturing

Simon Galle *

BI Norwegian Business School

May 2020

Abstract

This paper studies the dual impact of increased competition on aggregate output in a setting with both oligopolistic competition and financial constraints. In the absence of financial constraints, more competition unambiguously increases output by reducing markup levels, which increases aggregate capital. However, with financial constraints, stronger competition reduces the profitability of constrained firms and thereby slows down their rate of self-financed capital growth. Extensive reduced-form evidence confirms the theoretical predictions, including evidence from the pro-competitive impact of an industrial policy reform in India. In line with the theory, this reform reduces markup levels and dispersion, and slows capital growth. The quantitative analysis demonstrates that allocative efficiency declines with competition, but this negative effect on output is initially more than offset by a higher aggregate capital level due to lower markups. However, when firms have fixed operating costs, capital growth slows down drastically with competition, which eventually reduces aggregate capital. In this setting, less access to finance implies a lower optimal degree of competition.

*simon.galle@bi.no For invaluable advice, I am grateful to Ben Faber, Yuriy Gorodnichenko, Edward Miguel, Plamen Nenov and especially Andrés Rodríguez-Clare, and I am deeply thankful to Ishani Tewari and Hunt Allcott for sharing data. I also thank Juan-Pablo Atal, Pierre Bachas, Dorian Carloni, Fenella Carpena, Cécile Gaubert, Daniel Haanwinckel, Ben Handel, Alfonso Irarrázabal, Jeremy Magruder, Yusuf Mercan, Louis Raes, Marco Schwarz, Yury Yatsynovich and Moises Yi for many helpful discussions and suggestions; the International Growth Centre and the PEDL Initiative by CEPR and DFID for financial support to purchase the ASI data; and seminar participants at BI, Bocconi, CREI, ECARES, ECORES Summer School, Edinburgh, Erasmus Rotterdam, Gothenburg, HEC Lausanne, HEC Paris, IHEID Geneva, Johns Hopkins SAIS, KULeuven, NHH Bergen, PEDL-CEPR, Surrey, Tilburg, UC Berkeley, UCLA Anderson and University of Oslo for excellent comments. All errors are my own.

1 Introduction

Aggregate productivity is central to understanding why some countries are rich while others are poor. Since plant-level marginal productivities tend to be substantially more misaligned in poorer countries, resource misallocation, as proposed by Restuccia and Rogerson (2008), has become a prominent candidate for explaining differences in countries' aggregate productivity. While the potential factors contributing to misallocation are varied, the predominant view in the literature is that competition would be a beneficial force in reducing misallocation. After all, it is highly intuitive that competition will help shift resources from low-performing to high-performing plants, for instance by reducing markup levels and markup dispersion (Peters, 2016; Asturias, García-Santana, and Ramos, 2019), or by enhancing selection of high-productivity firms.

While the mechanisms driving competition's beneficial impact on aggregate productivity are undeniable, these beneficial mechanisms do not seem to cover the full story. Since limited access to finance is pervasive in developing countries (Levine, 2005), financially constrained firms often need to rely on retained earnings to finance their investments. Hence, profit levels determine how fast firms are able to save themselves out of their financially constrained position. Since competition reduces firms' profitability, it then also slows down investment for these firms. This, in turn, has negative implications for aggregate output.

This downside of competition may be especially salient in India, a large economy with strongly persistent levels of misallocation (Hsieh and Klenow, 2009; Bils, Klenow, and Ruane, 2020). Strikingly, most of that country's liberalization reforms, including an extensive licensing reform and a trade liberalization, had a null effect on the degree of allocative efficiency in its manufacturing sector (Bollard, Klenow, and Sharma, 2013). From the predominant perspective in the misallocation literature, this finding is puzzling. However, since even large Indian firms tend to be credit constrained (Banerjee and Duflo, 2014), it is important to take the interplay between competition and financial constraints into account in our understanding of the impact of competition on misallocation.

I develop a novel model to formally examine this interplay of competition and financial constraints. To allow for variation in competition and markup levels, the market structure is oligopolistically competitive as in Atkeson and Burstein (2008). In the absence of financial constraints, intensified competition decreases markups toward their lower bound, which increases aggregate capital accumulation and aggregate output. While this beneficial impact of competition on markup levels remains central in my framework, the introduction of financial constraints crucially leads to a second, harmful impact of competition on misallocation.

In each period in the model, a certain number of firms is newly born with a low level of initial assets and limited access to external finance. This limited access hampers their ability to grow their capital, in line with established stylized facts on young firms' financial constraints (Carreira and Silva, 2010; Fort, Haltiwanger, Jarmin, and Miranda, 2013), and leading to misallocation of capital (Midrigan and Xu, 2014). When access to external finance is limited, financially constrained firms rely on retained earnings to finance their investment. As a result, their rate of self-financed capital growth becomes a function of their profit level, which depends on their optimal markup. Increased competition, by reducing firms' markups, negatively affects their speed of capital growth. This way, competition amplifies "capital wedges" – the difference between constrained and unconstrained levels of capital – and thereby worsens capital misallocation. In the model, I derive the results on the dual impact of competition analytically: it reduces markup misallocation but amplifies capital misallocation.

After deriving these analytical results, I provide extensive reduced-form evidence in support of the theoretical predictions. To this end, I first leverage a natural experiment in India arising from the staggered implementation of an industrial policy change: the dereservation reform. Starting in 1997, this reform removed the investment ceilings imposed for the production of certain product categories, which led to the entry of new, larger firms in the production of the now dereserved product categories. Hence, the reform exposed incumbent plants to stiffer competition. Empirically, I examine the impact of the reform on incumbents' markups and on young plants' capital growth. I start by demonstrating in an event study that the dereservation reform leads to lower markups for incumbent plants, which confirms the pro-competitive impact of the reform. Moreover, markups for plants with an initially higher markup fall more than for plants with a lower initial markup, implying that the reform reduces markup dispersion. I also show that capital gowth for young plants slows down after the reform. Hence, the pro-competitive impact of the reform is in line with the theory: markup levels and markup dispersion fall, and so do capital growth rates for young plants.

To corroborate the external validity of the empirical analysis beyond the set of dereserved incumbent plants, I also examine capital growth for young plants on the full panel of plants. For the full panel, the measure of competition is the median markup across plants observed in the same state, sector and year. This median value is plausibly exogenous from the perspective of the individual plant. Again in line with the theoretical prediction, I document that a higher median markup is associated with faster capital growth for young plants. To further corroborate the theoretical mechanism, I explore how the impact of competition varies by a plant's degree of financial dependence. Employing the standard Rajan and Zingales (1998) measures, I find that plants in sectors with higher degrees of financial dependence exhibit a stronger sensitivity in their capital growth to the degree of competition.

After providing robust plant-level empirical support for the analytical predictions of the model, I turn to a quantitative analysis of increased competition in oligopolistic markets with financially constrained firms. I start by quantifying the theoretical predictions on how markups and capital growth fall as competition increases. Markups converge quickly toward their lower bound and the associated slowdown in capital growth is modest but economically meaningful. Importantly, I find that due to this slowdown, misallocation worsens with competition whenever firms have imperfect access to finance. In the baseline model however, the negative effect on aggregate output from lower allocative efficiency is more than offset by a higher aggregate capital level, which is driven by a reduction in markups. In this version of the model, it is optimal to take competition to its upper limit, even though the marginal benefit from competition becomes tiny when the initial number of firms is high.

In order to gain analytical tractability, the baseline model abstracts from two standard aspects of firm dynamics, namely fixed operating costs and love of variety. When I introduce these forces in the model, the theoretical predictions on competition's impact on markup levels and capital growth continue to hold. However, the presence of fixed costs implies that taking competition to its upper limit is no longer optimal. After all, even in the absence of financial constraints, the trade-off between love of variety and fixed costs leads to a finite optimal number of firms (Dixit and Stiglitz, 1977). In my setting this trade-off is enriched by the introduction of markup variation and financial constraints, and I find that the optimal number of firms falls as access to finance shrinks. When firm size shrinks due to increased competition, the share of the fixed cost in revenue rises, which reduces retained earnings and severly slows down the speed of capital growth. Beyond a certain threshold, this slowdown in capital growth becomes sufficiently large such that aggregate capital accumulation starts to fall with competition, which leads to a negative marginal effect of competition on aggregate output and consumption. Consequently, the presence of financial constraints leads to a lower optimal degree of competition than in the absence of these constraints.

Literature A closely related paper to mine is Itskhoki and Moll (2019), which also analyzes capital misallocation and examines how policy can affect investment through its impact on firm profitability. However, their main focus is on tax policy in a setting with perfect competition. In contrast, the key contribution of this paper is to examine the impact of competition on capital misallocation in an oligopolistic setting. Interestingly, this oligopolistic setting implies that the closed-form results from Moll (2014) no longer apply here. Still, by leveraging the logical relationships in the system of non-linear equations that describes the steady state, I am also able to derive analytical results on the interplay between the distribution of capital and the distribution of markups. More generally, my paper also relates to the macro-development literature on financial frictions, surveyed by Buera, Kaboski, and Shin (2015), and capital misallocation (Asker, Collard-Wexler, and De Loecker, 2014; Midrigan and Xu, 2014; Caggese and Pérez-Orive, 2017; Kehrig and Vincent, 2017)

Empirically, this paper focuses on testing the novel prediction of competition's negative impact on capital convergence, and I document robust support for this prediction across a series of plant-level tests. This evidence can help inform why misallocation has been persistent in India, despite several liberalization reforms. From that perspective, the paper complements existing studies on allocative efficiency in Indian manufacturing, including the analysis of markup misallocation (De Loecker, Goldberg, Khandelwal, and Pavcnik, 2016; Asturias et al., 2019), the role of financial constraints (Banerjee, Cole, and Duflo, 2005; Banerjee and Duflo, 2014), the impact of structural reforms (Aghion, Burgess, Redding, and Zilibotti, 2008; Sivadasan, 2009; Chari, 2011; Bollard et al., 2013; Alfaro and Chari, 2014), and the role of formal and informal institutions (Akcigit, Alp, and Peters, 2016; Boehm and Oberfield, 2018). Here, my paper is most closely related to the studies of the dereservation reform (García-Santana and Pijoan-Mas, 2014; Martin, Nataraj, and Harrison, 2017; Tewari and Wilde, 2017; Balasundharam, 2018; Boehm, Dhingra, and Morrow, 2019). These studies document the various beneficial impacts of this dereservation reform, while my analysis leverages the pro-competitive impact of the reform to test my model's predictions.

Taken together, the contribution of this paper is to develop a more nuanced understanding of the positive as well as the underexamined negative impact of competition on misallocation. These findings echo the ambiguous welfare impact of competition in other settings.[1] For instance, shielding an infant industry from competition may be beneficial if that industry has a latent comparative advantage. Importantly though, my model is more widely applicable than the infant

[1] In industrial organization, it is well-established that increasing competition can have both positive and negative effects on aggregate output or welfare. Negative effects can arise from business stealing (Mankiw and Whinston, 1986; Dhingra and Morrow, 2019) or by decreasing incentives to innovate (Gilbert, 2006; Aghion, Akcigit, and Howitt, 2014). In international trade, Foellmi and Oechslin (2016) show that increased competition due to trade can hamper credit access and thereby firm productivity, while Epifani and Gancia (2011) show that it can amplify cross-sectoral markup misallocation. In a more recent contribution, Jungherr and Strauss (2017) argue that higher market power is associated with higher growth in the Korean manufacturing sector.

industry argument, since financial constraints and market power are a general and robust feature of the data (Levine, 2005; De Loecker and Eeckhout, 2018), whereas the evidence on industries having a latent comparative advantage is mixed at best (Harrison and Rodríguez-Clare, 2010).

The next section presents the theory, Section 3 discusses the reduced-form results and Section 4 performs the quantitative analysis. Finally, Section 5 concludes.

2 Theory

2.1 Setup of the economy

Following Hottman, Redding, and Weinstein (2016), I assume that the economy has a continuum of sectors, and within each sector, there is a finite number of firms that produce differentiated goods. The final good Q_t^F is produced in a competitive market according to the following Cobb-Douglas production function:

$$\ln Q_t^F = \int_{s \in S} \phi_s \ln Q_{st} ds, \qquad \text{with} \int_{s \in S} \phi_s ds = 1, \tag{1}$$

where S is the measure of sectors and Q_{st} is a sector-level composite good for sector s in period t. An individual sector being atomistic relative to the macroeconomy will prove useful in the analytical derivation. Time is discrete. The standard price index P_t^F for the final good is $\ln P_t^F = \int_{s \in S} \phi_s \ln(P_{st}/\phi_s) ds$, where P_{st} is the price index for sector s. I choose the final good as the numeraire and set $P_t^F = 1$. A direct implication of this setup is that the optimal expenditure shares on goods for sector s are constant at $\phi_s = P_{st} Q_{st} / Q_t^F$.

The sector-level composite good for each sector, Q_{st}, is given by

$$Q_{st} = M_s^{\frac{1}{1-\sigma}} \left[\sum_{i=1}^{M_s} q_{ist}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where q_{ist} is consumption of the variety from firm i in sector s at time t, $\sigma > 1$ is the elasticity of substitution, and M_s is the exogenous number of firms in sector s. The fact that a sector's CES aggregate consists of a finite number of firms, as in Atkeson and Burstein (2008), implies that the intensity of competition is a function of that number of firms. The term $M_s^{\frac{1}{1-\sigma}}$ eliminates love of variety, as in Blanchard and Kiyotaki (1987). In the analysis below, this elimination of love of variety allows me to isolate the pro-competitive effects of changes in M_s. The inverse demand function and associated revenue function v_{ist} for variety i are then given by:

$$p_{ist}(q_{ist}, \mathbf{q_{-ist}}) = q_{ist}^{-1/\sigma} P_{st}(\mathbf{q_{st}})^{\frac{\sigma-1}{\sigma}} \left(\frac{\phi_s Q_t^F}{M_s} \right)^{1/\sigma}, \tag{3}$$

$$v_{ist}(q_{ist}, \mathbf{q_{-ist}}) = [q_{ist} P_{st}(\mathbf{q_{st}})]^{\frac{\sigma-1}{\sigma}} \left(\frac{\phi_s Q_t^F}{M_s} \right)^{1/\sigma}, \tag{4}$$

where $\mathbf{q_{st}} \equiv \{q_{ist}\}$ is the vector of all firms' quantities demanded in sector s, $\mathbf{q_{-ist}}$ the vector of all quantities for firm i's competitors, and the sectoral price index is

$$P_{st} = M_s^{\frac{1}{\sigma-1}} \left(\sum_{i=1}^{M_s} p_{ist}(\mathbf{q_{st}})^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{5}$$

In the oligopolistic setting under consideration, a firm's demand will become more inelastic as its market share m_{ist} increases:[2]

$$\varepsilon_{ist} \equiv -\frac{\partial q_{ist}}{\partial p_{ist}} \frac{p_{ist}}{q_{ist}} \quad \text{with} \quad \frac{\partial \varepsilon_{ist}}{\partial m_{ist}} < 0, \quad \text{and} \quad 1 \leq \varepsilon_{ist}(m_{ist}) < \sigma, \tag{6}$$

where the market share is defined as

$$m_{ist} \equiv \frac{v_{ist}}{\sum_{j=1}^{M_s} v_{jst}} = \frac{q_{ist}^{\frac{\sigma-1}{\sigma}}}{\sum_{j=1}^{M_s} q_{jst}^{\frac{\sigma-1}{\sigma}}}. \tag{7}$$

As a consequence, a firm's demand elasticity is a function of both its own demand and the demand of its competitors, $\varepsilon_{ist}(q_{ist}, \mathbf{q_{-ist}})$, which will matter below for firms' optimal markups.

Workers Workers are infinitely lived, supply labor inelastically, and each worker is hired at a wage w_t. These workers optimize their intertemporal utility over consumption c_{lt} of the final good:

$$E_r[U_{lt}] = \sum_{t=r}^{\infty} \beta^{t-r} E_r[c_{lt}], \tag{8}$$

where $0 < \beta < 1$ is workers' discount factor. Workers get paid at the end of each period and can use their earnings to consume at the start of the next one. These workers can lend to firms, using a one-period risk-free security b_{lt} with interest rate r_t^d; they may also receive a lump sum transfer ω_t, discussed below. Hence, their period-by-period budget constraint is

$$c_{lt} + b_{lt} \leq w_{t-1} + (1 + r_{t-1}^d) b_{lt-1} + \omega_t, \tag{9}$$

where consumption c_{lt} is constrained to be weakly positive. Workers' linear utility implies the following optimal choices for saving and consumption:

$$\begin{aligned} \left(r_t^d > \frac{1}{\beta} - 1\right) &\implies (b_{lt}^* > 0, c_{lt}^* = 0) \\ \left(r_t^d < \frac{1}{\beta} - 1\right) &\implies (b_{lt}^* < 0, c_{lt}^* > 0) \\ \left(r_t^d = \frac{1}{\beta} - 1\right) &\implies (c_{lt}^* \geq 0). \end{aligned} \tag{10}$$

Firms Each firm produces y_{ist}, the output for its variety, using capital k_{ist} and labor l_{ist} according to a Cobb-Douglas production function

$$y_{ist}(l_{ist}, k_{ist}) = k_{ist}^{\alpha} l_{ist}^{1-\alpha}. \tag{11}$$

Here, α is allowed to be sector specific, but for notational convenience I drop the subscript s.

At the start of each period, firms decide how much labor and capital to use in production. As mentioned above, wages are only paid at the end of the period, i.e. after revenue is realized.

[2] While the precise value of the demand elasticity will depend on the details of the oligopolistic market structure, the qualitative relation between market share and demand elasticity in equation (6) holds under both Bertrand and Cournot competition. In case firms engage in Cournot competition, their demand elasticity is $\varepsilon_{ist}(q_{ist}) = \left[\frac{1}{\sigma}(1 - m_{ist}) + m_{ist}\right]^{-1}$, and in case of Bertrand competition, it is $\varepsilon_{ist}(q_{ist}) = \sigma(1 - m_{ist}) + m_{ist}$ (see Atkeson and Burstein (2008), and Amiti, Itskhoki, and Konings (2016) for derivations.)

In contrast, investment in capital happens at the start of the period, when the firm owner also decides on consumption c_{ist} and her desired debt level d_{ist}. For this decision, the firm faces the following budget constraint

$$k_{ist} + c_{ist} \leq a_{ist-1} + d_{ist}, \tag{12}$$

where a_{ist-1} is net wealth accumulated in the previous period, and capital, consumption, wealth, and debt are all units of the numeraire, the final good. Importantly, the firm's borrowing is subject to a collateral constraint as in Moll (2014), which puts a limit on the firm's leverage ratio:

$$\frac{d_{ist}}{k_{ist}} \leq \lambda, \quad \text{with} \quad 0 \leq \lambda \leq 1. \tag{13}$$

At the end of the period, firms' capital has depreciated, and firm owners pay down their debt including an interest rate r_t^d. At this point, a firm owner's net real wealth is then

$$a_{ist} \equiv \pi_{ist}(l_{ist}, k_{ist}) + (1-\delta)k_{ist} - (1 + r_t^d)d_{ist}, \tag{14}$$

where $\pi_{ist}(l_{ist}, k_{ist}) \equiv v_{ist}(l_{ist}, k_{ist}) - w_t l_{ist}$ is revenue net of payments to labor. Each firm's wealth level is that firm's relevant state variable at the start of the next period, and as we will see below, in equilibrium it will always be strictly positive, assuming non-negative initial values.

Birth and death of firms In contrast to the infinitely lived workers, firm owners die with a positive probability $1 - \eta$. Such deaths occur after the end of the current period and before the start of the next. I assume that the value of firm owners' discount factor is $\beta_f = \beta/\eta$, which simplifies the analysis since it implies that they have the same intertemporal objective function as workers, namely equation (8). This way, workers and firm-owners have the same marginal utility for expected future consumption.

Since there are a finite number of firms in each sector, the law of large numbers does not hold within a sector. To make the analysis tractable, the death probabilities across firms are ex-ante identical, but not independent. Specifically, I assume that each period, a fixed number of firms equal to $(1 - \eta)M_s$ dies – with parameter values such that ηM_s is an integer. Before the start of each period a number $(1 - \eta)M_s$ of firms is newly born, which ensures that the number of firms in each sector is constant over time.[3] The wealth of the deceased firms, which will always be positive in steady state equilibrium, is distributed partly as capital endowments to newborn firms and partly as lump sum transfers to workers. The reasons why wealth at the end of any period is positive, as well as the details of the wealth allocation process are discussed below. Importantly, if both λ and newborn firms' initial wealth are sufficiently low, younger firms will need to grow their capital over time, which leads to dispersion in marginal products of capital.

[3] The death process is further defined across $T^{\bar{M}} + 1$ different age bins, with $T^{\bar{M}}$ specified below. Among all firms older than $T^{\bar{M}}$, the death process is such that a fixed share $1 - \eta$ dies each period. In addition, among all firms weakly younger than $T^{\bar{M}}$, there is also a fixed share $1-\eta$ of each specific age that dies. I make this assumption to allow for a stable capital distribution over time, which requires a constant number of firms in all constrained age bins, and a fixed number of unconstrained firms (see Lemma 1). Finally, define $T^{\bar{M}}$ such that firms are unconstrained after $T^{\bar{M}}$ periods when $M_s = \bar{M}$. Then, since capital growth slows down with M_s (see Proposition 1), this ensures a stable capital distribution for all values $M_s \leq \bar{M}$. This is without loss of generality, since $\bar{M}$ can be set arbitrarily high. The only constraint is that parameter values should be such that the number of firms dying in each bin is an integer.

Marginal cost functions The above intertemporal setup implies that firms' opportunity cost of inputs, in terms of utility from consumption in period $t+1$, is w_t for labor and $r^k \equiv \frac{1}{\beta} + \delta - 1$ for capital. Given these input costs, standard cost minimization for Cobb-Douglas production functions implies that unconstrained firms have the following factor demands

$$k^u(y_{ist}) = \left(\frac{w_t}{r^k} \frac{\alpha}{1-\alpha} \right)^{1-\alpha} y_{ist}, \tag{15}$$

$$l^u(y_{ist}) = \left(\frac{r^k}{w_t} \frac{1-\alpha}{\alpha} \right)^{\alpha} y_{ist}, \tag{16}$$

which result in the following constant marginal cost for an unconstrained firm:

$$MC^u_{st} = \left(\frac{r^k}{\alpha} \right)^{\alpha} \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha}. \tag{17}$$

The budget constraint in equation (12) and the collateral constraint in (13) imply that a firm's maximum capital level is $k^c_{ist} \equiv a_{ist-1}/(1-\lambda)$. For any desired quantity $\bar{y}_{ist}$ where a firm is constrained – i.e. where $k^u(\bar{y}_{ist}) > k^c_{ist}$ – a firm will set its capital at its maximum (k^c_{ist}), since a lower level of capital would imply a higher loss in terms of utility from consumption. When $k_{ist} = k^c_{ist}$, a firm can only adjust its labor input at the margin and therefore its total variable costs are $w_t l(\bar{y}_{ist})$. This implies the following marginal cost function:

$$MC^c_{st}(\bar{y}_{ist}, k^c_{ist}) = \frac{w_t}{(1-\alpha)} \left(\frac{\bar{y}_{ist}}{k^c_{ist}} \right)^{\frac{\alpha}{1-\alpha}}, \tag{18}$$

which is increasing in $\bar{y}_{ist}$, and is strictly higher than MC^u_{st} for $k^u(\bar{y}_{ist}) > k^c_{ist}$. Combining equations (17) and (18), and recalling that k^c_{ist} is a function of a_{ist-1}, a firm's marginal cost is a function of its output level and wealth: $MC_{ist}(y_{ist}, a_{ist-1})$.

Market structure The firms play an infinitely repeated quantity-setting game. Here, the state of a firm's competitors can be summarized by $D^s(\mathbf{a_{-ist-1}})$, the distribution of wealth for all firms in sector s excluding firm i, while $D^s(\mathbf{a_{ist-1}})$ denotes the distribution of wealth for *all* firms in industry s, i.e. the state of the industry. Firms' strategies formulate actions conditional on a firm's own state, and the state of its competitors.[4] Here, I define an industry equilibrium as a set of strategies for all firms in sector s that constitute a Nash equilibrium, given a specific path for the macroeconomy $\mathbf{F_t} \equiv \{w_t, Q^F_t, r^d_t\}$.

Since the game is infinitely repeated, there are many industry equilibria. In my analysis, I focus on a natural benchmark equilibrium, namely the repetition of the static game.[5]

[4]Formally, a strategy of a firm consists of a set of decision rules for capital, labor and debt, valid for all current and future periods, that are conditional on the firm's own state a_{ist-1}, the state of its competitors $D^s(\mathbf{a_{-ist-1}})$, the history of the game, and the state of the macroeconomy, summarized by $\mathbf{F_t} \equiv \{w_t, Q^F_t, r^d_t\}$. In each sector, each firm then chooses the strategy that maximizes its present value of consumption, conditional on the strategies of its competitors and subject to their budget and collateral constraint in (12) and (13). This optimization implies that the budget constraint is satisfied with equality, which means that decisions for capital and debt imply a decision for consumption. In addition, decisions on labor and capital imply a decision for output.

[5]In a standard, one-shot Cournot game, firms' production yields a price that results in an optimal markup over their marginal cost, given their residual demand function. In the case without financial constraints ($\lambda = 1$), it is straightforward to show that a repetition of the action in the one-shot game is a subgame perfect Nash equilibrium. However, the folk theorem logic implies that more collusive outcomes can also be subgame perfect (Friedman, 1971). In a setting with financial constraints ($\lambda < 1$), the analysis of subgame perfect equilibria becomes more complex however, for instance because trigger strategies can lead $D^s(\mathbf{a_{ist-1}})$, the state of the industry, to become non-stationary. As a result, subgame perfect equilibria are challenging to characterize analytically in the current setting with financial constraints.

Assumption 1. *On the equilibrium path, firms' output responses, or reaction functions, to the output of their competitors are implicitly defined by:*

$$\frac{\varepsilon_{ist}(y_{ist}, \mathbf{y_{-ist}})}{\varepsilon_{ist}(y_{ist}, \mathbf{y_{-ist}}) - 1} = \frac{p_{ist}(y_{ist}, \mathbf{y_{-ist}})}{MC_{ist}(y_{ist}, a_{ist-1})}. \tag{19}$$

It is relatively straightforward to show that, for certain trigger strategies and given a condition on the discount factor β, these reaction functions constitute a standard Nash equilibrium. In this equilibrium, each firm's decisions on output, capital and labor, depend on their own state, the decisions of their competitors and the state of the macroeconomy:

$$k^*_{ist}(a_{ist}, D^s(\mathbf{a_{-ist-1}}), \mathbf{F_t}),$$
$$l^*_{ist}(a_{ist}, D^s(\mathbf{a_{-ist-1}}), \mathbf{F_t}),$$
$$d^*_{ist}(a_{ist}, D^s(\mathbf{a_{-ist-1}}), \mathbf{F_t}).$$

As a consequence, the state of the sector $D^s(\mathbf{a_{ist-1}})$ and the state of the macroeconomy together determine the joint distribution of capital and labor, denoted by $H^s(k_{ist}, l_{ist})$.

2.2 Steady state equilibria

Definition. *A **steady state equilibrium** consists of, first, stable industry equilibria for all sectors, where within each sector the distribution of wealth and the joint distribution of capital and labor are stable:*

$$D^s(a_{ist-1}) = D^s(a),$$
$$H^s(k_{ist}, l_{ist}) = H^s(k, l).$$

Second, workers' decision rules for saving and consumption described in (10) *that satisfy the budget constraint in equation* (9) *with equality. Third, a stable macroeconomic state* $\mathbf{F}$*: a wage* w*, an interest rate* r^d*, and total output of the final good* Q^F *such that the labor market clears in every period:*

$$L = \int_{s \in S} \sum_{i=1}^{M_s} l^*_{ist}(a_{ist-1}, D^s(\mathbf{a_{-ist-1}}), \mathbf{F}) ds, \tag{20}$$

and the debt market clears given decisions about investment and consumption:

$$\int_{l \in L} b^*_{lt}(\mathbf{F}) dl = \int_{s \in S} \sum_{i=1}^{M_s} d^*_{ist}(a_{ist-1}, D^s(\mathbf{a_{-ist-1}}), \mathbf{F}) ds. \tag{21}$$

Note that by Walras' law, since all firms and all workers satisfy their budget constraints with equality, when labor and debt markets clear, the goods market also clears. In a steady state equilibrium, the interest rate will be $r^d = \frac{1}{\beta} - 1$. This is because a higher interest rate would over time lead to an excess supply of saving, due to wealth accumulation by workers and unconstrained firms; and a lower interest rate implies excess demand for borrowing by workers and unconstrained firms.

Given this interest rate and the firms' equilibrium behavior, firms are always able to repay their debt at the end of the period. First, note that marginal cost is weakly higher than average cost, and strictly so when the firm is constrained. In addition, since $\sigma/(\sigma-1)$ is a lower bound on the markup, markups are always above unity. Together, this implies that firms can always cover the total opportunity cost of capital and labor. Then, because debt is always weakly lower than capital due to the collateral constraint in (13), firms always repay their full debt at the end of the period, and therefore their wealth is always positive.

Distribution of capital The equilibrium for the reaction functions in Assumption (1), entails a certain output level y_{ist} for each firm. As explained when deriving the marginal cost function, firms can either be constrained or unconstrained when setting their capital input to reach this output level, and constrained firms set their capital level at $k^c_{ist} \equiv a_{ist-1}/(1-\lambda)$. The unconstrained capital level within each sector is stable over time in a steady state equilibrium, and denoted by k^u_s.

As mentioned above, all newborn firms receive a starting level of wealth from deceased firms, and I denote this by a_{s-1}. Importantly, I express this endowment as a share of the unconstrained capital level:

$$a_{s-1} = \zeta_s k^u_s,$$

with $0 < \zeta_s \leq 1$. The remaining wealth of the deceased firms that is not redistributed to newborn firms is divided across all workers in a transfer ω.

In summary, $a_{s-1} = \zeta_s k^u_s$ is the initial wealth of a newborn firm, and equation (14) describes the equation of motion for wealth at the end of each period. Together with constrained firms' capital being equal to $k^c_{ist} \equiv a_{ist-1}/(1-\lambda)$, this implies that one can solve exactly for the wealth level of each constrained firm as a function of its age τ, as in the following Lemma.

Lemma 1. *In steady state, the joint distribution of capital and productivity within a sector is as follows:*

- *unconstrained firms set* $k_{ist} = k^u_s$
- *constrained firms have* $k_{ist} = \frac{G_{s\tau-1}a_{s-1}}{1-\lambda}$, *where* $a_{s-1} = \zeta_s k^u_s$, *and where* $G_{s\tau-1}$ *is the cumulative growth rate of wealth over the past* $\tau-1$ *periods for a firm of age* τ *at time t:*

$$G_{s\tau-1} \equiv \frac{a_{s\tau-1}}{a_{s-1}} = \Pi_{v=0}^{\tau-1}\left(\frac{\pi_{sv}}{a_{sv-1}} + \frac{1-\delta-\lambda-r^d}{1-\lambda}\right).$$

Naturally, if firms are constrained for T periods, there are $T+1$ different capital levels in a sector. As a consequence, to allow for a steady state, the death process needs to be such that a constant number of firms of each capital type dies (see footnote 3).

2.3 No capital constraints

To establish a benchmark, I first consider the case where $\lambda = 1$ and firms therefore face no limit on their capital levels. Since all firms are unconstrained, have an identical constant marginal cost, and have symmetric reaction functions in a sector, they set the same level of output, labor and capital. Defining aggregate capital and labor in a sector as $K_s \equiv \sum_i k_{ist}$ and $L_s \equiv \sum_i l_{ist}$, each firm has the same share of labor and capital in a sector: $k_{ist}/K_{st} = l_{ist}/L_{st} = 1/M_s$. As a result, marginal products are perfectly equalized across firms, such that there is no misallocation of resources. This is reflected in a sectoral total factor productivity of one:

$$TFP_{st} \equiv \frac{Q_{st}}{K_{st}^{\alpha} L_{st}^{1-\alpha}} = 1.$$

Hence, TFP_{st} is invariable to M_s when there are no capital constraints. However, comparing across industry equilibria while holding the state of the macro-economy constant, other variables do change with M_s. To start, recall that a firm's price is a markup over marginal cost: $p_{ist} = MC_{ist}^{u} \mu_{ist}(m_{ist})$, with $\mu_{ist}(m_{ist}) \equiv \varepsilon_{ist}(m_{ist}) / \left(\varepsilon_{ist}(m_{ist}) - 1\right)$. Since market shares m_{ist} fall monotonically as M_s increases, the demand elasicity increases and firms' markups and prices decrease with M_s. Since firms are symmetric, the sectoral price index is identical to firms' prices ($P_{st} = p_{ist}$), and this price index also falls with M_s. As a result, sectoral output, in equilibrium equal to $Q_{st} = \phi_s Q^F / P_{st}$, increases with M_s.

In this analysis, where macroeconomic variables are held constant, sectoral output increases due to an increase in sectoral capital and labor. In the quantitative analysis, I will examine the general equilibrium impact of competition by increasing M_s in all sectors. While aggregate labor is then necessarily fixed, I show that the associated reduction in markups increases output by pushing up aggregate capital accumulation.

2.4 Comparative statics on the degree of competition

Now I move to a comparison of capital growth rates and markup levels across industry equilibria when firms face constraints on their capital ($\lambda < 1$). I continue to hold macroeconomic variables constant. To set up the analysis, I consider industry equilibria under two different values for the number of firms: $M_s \neq M_s'$, and all equilibrium values under M_s' are denoted with a prime. Initially, I will be agnostic about whether $M_s > M_s'$, and I start instead by supposing, without loss of generality, that the unconstrained firms' market share is higher in the former equilibrium ($m_s^u > m_s^{u'}$). I then examine the logical implications of that supposition on other firms' market shares and capital growth rates. Those logical implications, summarized in Lemma 2, will in turn imply that $M_s < M_s'$, which will allow me to conduct comparative statics across equilibria with different numbers of firms. Note that from now on in this theory section, the notation for different types of firms follows Lemma 1; only constrained firms are denoted with the subscript τ, indicating their "age bin" τ, and I drop the superscript c. I also focus exclusively on equilibria where at least the newborn firms are constrained, which is guaranteed for sufficiently low values of ζ_s and λ.

Lemma 2. *If the market share of the unconstrained firms is higher in one industry equilibrium compared to another, then the market share of all constrained firms is also higher, and so is their capital growth rate*

$$(m_s^u > m_s^{u'}) \implies \forall \tau \geq 0 : (m_{s\tau} > m_{s\tau}') \wedge (G_{s\tau} > G_{s\tau}') \tag{22}$$

The proof in Appendix Section A.1 proceeds by induction. There, I first demonstrate that

$$(m_s^u > m_s^{u'}) \implies ((m_{s0} > m_{s0}') \wedge (G_{s0} > G_{s0}')) \,.$$

The implication on market shares follows from the fact that newborn firms inherit a constant share of capital from unconstrained firms. Suppose then to the contrary that unconstrained firms have a higher market share while newborn firms have a lower market share in the former equilibrium. Comparing relative equilibrium demand for the newborn and unconstrained firms, this would require the newborn firms to have a higher price. This higher price could result from a higher marginal cost or a higher markup, but given the constant share of capital newborn firms inherit, both the higher markup or marginal cost would require a higher level of output. This results in a contradiction and hence newborn firms' market shares are also higher in the former equilibrium. Next, I show that since the newborn firms have a higher markup – as associated with their higher market share – their capital growth rate is also higher. This establishes the second component of the implication.

Then I demonstrate an inductive step that higher market shares for unconstrained firms and higher capital growth rates for constrained firms of age $\tau - 1$ result in higher market shares and capital growth rates for constrained firms of age τ:

$$\Big((m_s^u > m_s^{u'}) \wedge (G_{s\tau-1} > G'_{s\tau-1})\Big) \implies ((m_{s\tau} > m'_{s\tau}) \wedge (G_{s\tau} > G'_{s\tau}))\,.$$

Here, the proof follows a similar logic as before for each of the components of the implication. Since unconstrained firms have a higher market share, and firms in bin τ start off with a higher capital share, the constrained firms in bin τ having a lower market share results in a contradiction. Then, because they have a higher market share and therefore also a higher markup, their capital growth rate is increased as well.

Lemma 2 implies that the market shares of all types of firms – i.e. all unconstrained firms and constrained firms in any bin τ – jointly increase or decrease across industry equilibria.[6] Hence, when we have $(m_s^u > m_s^{u'})$, market shares for all types of firms are higher. When market shares for all types of firms are elevated, this implies that there are fewer firms in the industry, and therefore $(m_s^u > m_s^{u'}) \implies (M_s < M'_s)$. The converse also holds, since $(M_s < M'_s)$ implies that the market share of at least one type of firm needs to strictly decrease. Since all market shares increase and decrease together (see footnote 6), it follows that:

$$(M_s < M'_s) \implies (m_s^u > m_s^{u'}).$$

In combination with Lemma 2, this directly implies that when the number of firms falls, the market share of all types of firms increases:

$$(M_s < M'_s) \implies \Big((m_s^u > m_s^{u'}) \wedge (\forall \tau \geq 0 : m_{s\tau} > m'_{s\tau})\Big) \tag{23}$$

This result, combined with the monotonically increasing relationship between market shares and markups, implied by equations (6) and (19), directly implies that markup levels for all types of firms fall as the number of firms increases. Moreover, together with Lemma 2, it implies that capital growth rates of constrained firms fall as well when the number of firms increases. Finally, Appendix Section A.2 demonstrates that markup dispersion also falls with the number of firms. This is intuitive, since as $M_s \to \infty$, all markups converge to $\sigma/(\sigma - 1)$, the markup under monopolistic competition. All the above together leads to the following proposition:

Proposition 1. *For any $M'_s > M_s$, unconstrained firms u, and constrained firms in any bin $\tau \geq 0$:*

- *Markup levels fall with M_s:*

$$\mu_s^{u'} < \mu_s^u \;\; ; \;\; \mu'_{s\tau} < \mu_{s\tau}$$

[6] Naturally, the statement $(m_s^u > m_s^{u'}) \wedge (\exists \tau \geq 0 : m_{s\tau} \leq m'_{s\tau})$, contradicts Lemma 2.

- *Markup dispersion falls with M_s:*

$$\frac{\mu_s^{u\prime}}{\mu_{s0}'} < \frac{\mu_s^u}{\mu_{s0}} \quad ; \quad \frac{\mu_{s\tau}'}{\mu_{s0}'} \leq \frac{\mu_{s\tau}}{\mu_{s0}}$$

- *Capital growth rates for all financially constrained firms fall with M_s:*

$$G_{s\tau}' < G_{s\tau}.$$

The results in Proposition 1 are highly intuitive, but not obvious. After all, since the capital growth rate $G_{s\tau}$ falls with M_s, it could have been the case that the market shares of the unconstrained firms increase with M_s. The above analysis verifies that this is not the case, and that both capital growth rates and markup levels fall monotonically with the number of firms.

These results can also have important welfare implications, in particular for understanding the gains from taking competition to its upper limit. In the case with no capital constraints, setting $M_s \to \infty$ unambiguously increased output. In contrast, Proposition 1 entails that $\frac{k_{s\tau}}{k_{s0}}$ falls with M_s, such that the wedge between unconstrained and constrained capital levels deepens. In the simulation exercise in Section 4, I will look at the broader implications of Proposition 1 on TFP and output. Before that, I document the empirical relevance of the channels in Proposition 1.

3 Reduced-form evidence for the theoretical prediction

To empirically test the model predictions on the impact of competition at the plant level, I start by leveraging natural variation in competition arising from India's dereservation reform. In line with the theory, among incumbent plants exposed to this pro-competitive reform, markup levels and markup dispersion fall after the reform, and so does the capital growth rate of young plants. Additionally, to strengthen the external validity of the empirical evidence beyond the sample of dereserved incumbent plants, I document a negative association between competition and capital growth for young plants in the full panel of Indian plants.

3.1 Panel data on Indian plants

I employ data on manufacturing establishments, or plants, from the Indian Annual Survey of Industries (ASI), for the period 1990-2011. The ASI provides data by accounting year, which starts on April 1st. I refer to each accounting year by the calendar year on which it starts. The ASI sampling scheme consists of two components. The first component is a census of all manufacturing establishments located in a small number of specific geographic areas, or above a certain size threshold. For almost all years in my dataset, the size threshold for the census component is an employment level of 100 workers. The only exceptions are the years 1997 until 1999, when this threshold is 200 workers instead. The second component of the sampling scheme includes, with a certain probability, each formally registered establishment that is not part of the census component. All establishments with more than 20 workers, or 10 workers if the establishment uses electricity, are required to be formally registered. For my analysis, I restrict the sample to manufacturing sector plants that are operational and have non-missing, positive values for the logarithm of three critical variables in the analysis, namely revenue, capital, and labor cost. I also ensure that definitions of geographical units, i.e. states or union territories, are consistent over time. Appendix B gives a full overview of the data cleaning procedure.

Central to my analysis are the establishment identifiers, used to construct the panel for the entire 1990-2011 period. To construct the panel, I use the establishment identifiers provided by the Indian Statistical Office for all years from 1998 onward. For the years prior to 1998, I use the establishment identifiers employed by Allcott, Collard-Wexler, and O'Connell (2016), who gained access to these identifiers while working in India.[7]

For analyzing the industrial policy reform, I use data on the timing of dereservation for each dereserved product, which is available on the website of the Indian Ministry of Micro, Small and Medium Enterprises.[8] Importantly, the ministry sets the time of dereservation by SSI product code. To match these SSI product codes to the product classification in the ASI data, I use the concordance available from Martin et al. (2017). Appendix Section B.2 provides more detail on the construction of this concordance.

3.2 Background on the dereservation reform

The dereservation reform consists of the staggered removal of the small-scale industry (SSI) reservation policy. This reservation policy mandated that only industrial undertakings below a certain ceiling of accumulated investment - 10 million Rupees at historical cost in 1999 (Martin et al., 2017) - were allowed to produce products within certain product categories.[9] As such, this policy was one of the most important aspects of India's economic agenda of promoting small-scale industries (Mohan, 2002). This agenda was launched in the 1950s in an attempt to promote social equity and to boost employment growth by stimulating intensive use of labor in manufacturing. The reservation policy itself was introduced in the Third Five Year Plan (1961-1966). In 1996, before the start of dereservation, more than 1000 product categories were reserved for SSI. These product categories encompassed many different industries, such as chemicals, car parts, electronics, food and textiles. In total, reserved products constituted around 12% of Indian manufacturing output (Tewari and Wilde, 2017).

Dereservation started in 1997, several years after a first wave of liberalization policies in the early 1990s (Bollard et al., 2013). According to Martin et al. (2017), stiffer foreign competition following the trade liberalization and increased complexity of industrial production convinced policymakers to gradually start abandoning the reservation policy. The actual decision to dereserve a particular product is only taken after a series of meetings between relevant stakeholders, review up a chain of bureaucrats, and final approval by the central government minister (Tewari and Wilde, 2017). Appendix Figure C.1 provides an overview of the timing of dereservation. The process of dereservation clearly peaks between 2002 and 2008. By 2015, no products are reserved anymore.

I focus on "incumbent" plants, and I define a plant as incumbent if it produces at least one reserved product prior to dereservation. For these incumbent plants, I define their year of dereservation as the accounting year in which a first product of theirs is dereserved.[10] Importantly, as

[7] I thank Hunt Allcott for generously making these panel identifiers available.

[8] This list is available at `http://dcmsme.gov.in/publications/circulars/newcir.htm#RESERVED`, as retrieved on February 15, 2019.

[9] At the time of reservation, an exception was made for large industrial undertakings already producing the product. These undertakings were allowed to continue production, but with output capped at existing levels.

[10] Since the accounting years in the ASI data start on April 1st, the dereservation year for products dereserved between January 1st and March 31st, is the accounting year starting in the previous calendar year. For instance, products

described by Martin et al. (2017), the large majority of plants who produce reserved products - 89.5% of them in my sample - produce only one such product. Moreover, 93.7% of the plants make products that are all dereserved in the same accounting year. Appendix Table C.1 presents summary statistics for the panel of incumbent plants used in the estimation of capital growth below. The average incumbent in my sample is 18 years old and has 190 workers.

Since the decision to dereserve a product is made by policymakers in consultation with stakeholders, one may be worried about the threats this political process poses to causal identification of the reform's effects. However, several reasons indicate dereservation was implemented in a semi-random manner. First, all reserved products are eventually dereserved, which precludes selection into being dereserved or not. Second, there is no evidence for selective timing of dereservation. To start, Tewari and Wilde (2017) demonstrate that there is considerable variation in the timing of dereservation for strongly related product categories, e.g., different types of vegetable oils. As products within these narrow product categories arguably share similar demand and supply characteristics, this limits the scope for a structural explanation of the timing of dereservation. More importantly, there are no pre-trends associated with the timing of dereservation for plant-level revenue, capital, labor or labor cost (see Appendix Figure C.2). This evidence is in line with a result from Martin, Nataraj, and Harrison (2014) on the absence of pre-trends associated with dereservation timing for plant-level employment. Taking everything together, this evidence lends credible support to the timing of dereservation being semi-random.

3.3 Dereservation as a pro-competitive shock

What are the main effects of deservation? This policy reform implies both the removal of a size restriction, and an increase in the degree of competition. The latter effect arises from incumbent plants being allowed to grow their capital stock, and in particular from larger firms entering the previously reserved product categories. Clearly, the removal of the size restriction has direct positive effects on allocative efficiency (Guner, Ventura, and Xu, 2008), the welfare impacts of which are analyzed by García-Santana and Pijoan-Mas (2014). In this paper however, I focus instead on the pro-competitive aspect of the reform to test the theoretical predictions of my model.[11]

The pro-competitive shock due to dereservation is large. As a first, strong indication of the increase in competition, note that over the entire sample period there are only 38,592 unique incumbent plants in my dataset, whereas 113,967 unique plants enter the previously reserved product categories after dereservation. Martin et al. (2017) provide a detailed analysis of the pro-competitive impact of dereservation, demonstrating that for entrant plants, employment, output and capital grow substantially after dereservation, whereas the average incumbent plant shrinks on these dimensions.

I start by investigating how dereservation affects the markups of incumbent plants. I measure μ_{it}, the markup for plant i in year t, as in De Loecker and Warzynski (2012)'s application of Hall (1986)'s insight:

dereserved on February 3d 1999, have the 1998-1999 accounting year as their year of dereservation.

[11] Critically then, I am not performing a complete analysis of the impact of the reform on allocative efficiency in the manufacturing sector. After all, this would require bringing the additional complication of size restrictions into the model. Since the steady state of the model is already described by a system of non-linear equations, it is unclear if analytical results would still be available in that case. Instead of performing a welfare analysis of the dereservation reform, I am interested in examining if the analytical predictions of my model are borne out by the empirical reality. To that end, I exploit the natural variation in competition for incumbent plants arising from the dereservation reform.

$$\mu_{it} = \alpha_i^L \frac{S_{it}}{w_{it} L_{it}} \tag{24}$$

where α_i^L is the plant-level elasticity of value added with respect to labor, and $\frac{w_{it}L_{it}}{S_{it}}$ is labor's share of revenue S_{it}.[12] Intuitively, assuming a constant output elasticity of labor, a higher labor share implies a lower markup in this measure. This expression for the markup rests on the assumptions of a Cobb-Douglas production function, cost minimization, and having labor as a variable input.[13] The latter assumption appears plausible in my data, since after dereservation the number of employees immediately starts falling significantly and substantially for incumbent plants (see Panel c of Appendix Figure C.2.)

Using this markup measure, I run the following event-study on dereservation:

$$\ln \mu_{it} = \gamma_i + \nu_t + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{it} + \tau] + \varepsilon_{it} \tag{25}$$

Here, ν_t is a year fixed effect, and γ_i is a plant fixed-effect that absorbs α_i^L such that I allow for maximal cross-plant heterogeneity in this output elasticity. The year in which a plant's first product is dereserved is denoted by e_{it}, and I bin up the end points and normalize $\beta_{-1} = 0$. For the purpose of this event study, I restrict attention to a balanced sample of incumbent plants which are observed at least three years before and after they were dereserved. An incumbent plant is any plant that produced a reserved product prior to dereservation. Since dereservation status is determined at the product level, I cluster standard errors at that level.

I find that on average, markups indeed fall due to the dereservation reform (see Figure 1, Panel a). The initial decline is modest, but eventually the average markup declines by 0.08 log points (p=0.086). This impact of dereservation is in line with the theoretical prediction that markup levels fall when competition increases, and the economic magnitude of the impact of the reform is substantial. Note also that there is no pre-trend for markups prior to dereservation.[14]

In addition to decreasing the average markup, dereservation also reduces markup dispersion. Recall that the model predicts that as the degree of competition increases, all markups converge to a lower bound. To test this prediction, I split the set of incumbent plants into two subsets depending on whether, before dereservation, a plant's markup is above or below the median initial markup.[15] I find that plants that have higher markups in the periods before dereservation exhibit a stronger average decline in their markup after dereservation, as predicted by the theory. More specifically, for the subset of plants with below-median initial markup, dereservation appears to have no effect on markup levels (see Figure 1, Panel (c)). In contrast, plants with above-median initial markups, experience a strong decline in their markup. After three years, their average

[12] In my data, the distribution of the inverse of the labor share of revenue appears roughly lognormal (see Appendix Figure C.3).

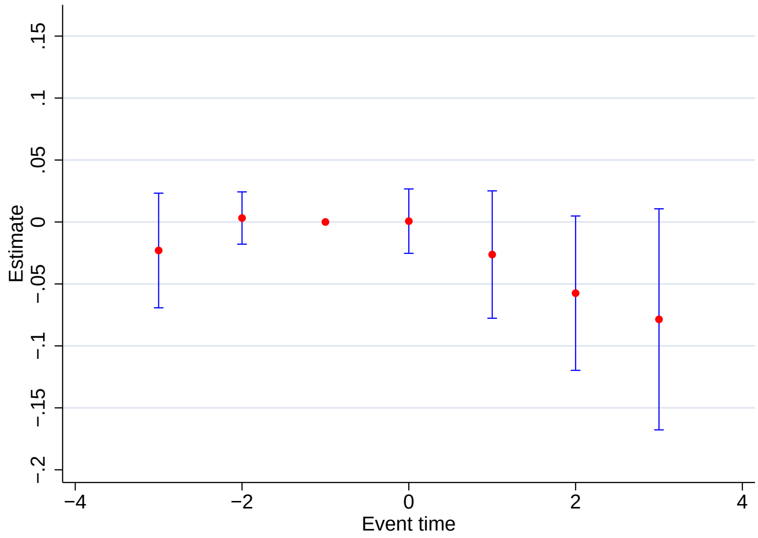
[13] Given a Cobb-Douglas production function, the Lagrangian of the cost-minimization problem with variable labor input and a predetermined capital level is $\min_{l_{it}} \mathcal{L}_{it} = w_{it} l_{it} + \lambda_{it}(Y_{it} - a_{it} k_{it}^{\alpha_i^K} l_{it}^{\alpha_i^L})$. Optimization sets: $w_{it} = \lambda_{it} \alpha_i^L \frac{a_{it} k_{it}^{\alpha_i^K} l_{it}^{\alpha_i^L}}{l_{it}}$, and therefore $\frac{p_{it}}{\lambda_{it}} = \alpha_i^L \frac{p_{it} a_{it} k_{it}^{\alpha_i^K} l_{it}^{\alpha_i^L}}{w_{it} l_{it}}$. Since λ_{it} is the marginal cost of output, $\frac{p_{it}}{\lambda_{it}} = \mu_{it} = \alpha_i^L \frac{p_{it} y_{it}}{w_{it} l_{it}}$.

[14] To further corroborate the absence of a pre-trend, I re-estimate the event study over a longer time horizon in Appendix Figure C.4. The results are qualitatively very similar for the plants below or above the median initial markup. The longer time horizon implies that the number of observations included in the balanced panel shrinks from 20,937 for the analysis in Figure 1 to 13,522 in Appendix Figure C.4. Together with the heterogeneity across plants with high or low initial markup, this reduced sample size renders the estimates in Panel (a) less statistically significant.

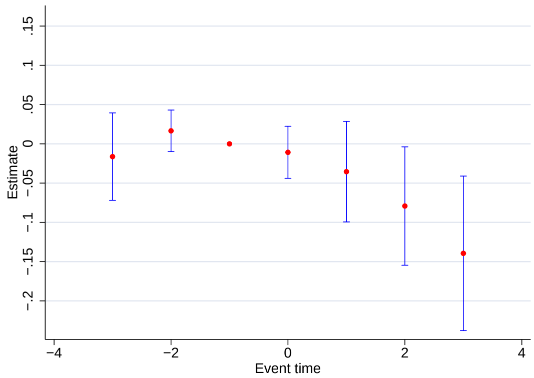
[15] The initial markup is averaged over event times $\tau = -3$ and $\tau = -2$, and the median initial markup is then set after taking out sector and year fixed effects.

Figure 1: Event study for the impact of dereservation on markups

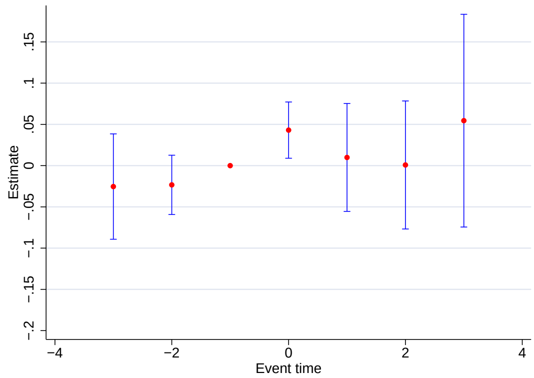
(a) Full sample of incumbents



(b) Incumbents with above-median initial markup



(c) Incumbents with below-median initial markup



The figure displays the coefficients and 95% confidence intervals for the β_τ coefficients from the following event-study regression: $\ln \mu_{it} = \gamma_i + \nu_t + \sum_{\tau=-3}^{3} \beta_\tau 1[t = e_{it} + \tau] + \varepsilon_{it}$, where γ_i is a plant fixed effect and ν_t is a year fixed effect. I define the time at which a plant's first product of plant i is dereserved as e_{it}, and I impose the normalization that $\beta_{-1} = 0$. Since dereservation status is defined at the product-level, I also cluster standard errors at that level. I restrict the sample to a balanced sample of incumbent plants that are observed at least three years before and after they are dereserved. Panel (a) shows results for the full sample of incumbent plants. Panel (b) displays results for plants with initial markups weakly above the median initial markup, and Panel (c) for the other plants. The initial markup is an average over event times $\tau = -3$ and $\tau = -2$, and the median initial markup is then set after taking out sector and year fixed effects. Since the ASI's product classification changes after 2010, this analysis is performed on plants who are dereserved before 2010. Additionally, Appendix Figure C.6 implements a robustness check by dropping plants dereserved before 1999, since product coverage in the ASI is incomplete before 1999.

markup is 0.14 log points lower (see Figure 1, Panel (b)).[16]

[16]What is driving this drop in markups for this latter group of plants? Interestingly, while these plants shed labor, if anything their labor cost seems to increase (see Appendix Figure C.5 Panels a and b). Combined with the downward pressure on revenue (Panel c), this leads to lower markups.

3.4 Capital growth

Now I test the model prediction on how capital growth for young plants falls when competition is higher. I measure capital growth as $g(k_{irst}) = \ln(k_{irst+1}/k_{irst})$, for firm i in state (or region) r and sector s. Here, capital is measured as the book value of assets, which is observed both at the start of year t, and at the end. The latter value is used as measure for k_{irst+1}. I deflate the book value of capital using the capital deflator from the Indian Handbook of Industrial Statistics.

To examine the effect of dereservation on capital growth for young plants, I estimate the following specification on a sample of incumbent plants:

$$\begin{aligned} g(k_{irst}) = \gamma_r + \delta_s + \zeta_t + \beta_1 young_{irst} + \beta_2 Deres_{irst-1} \\ + \beta_3 Deres_{irst-1} * young_{irst} + \varepsilon_{irst}, \end{aligned} \tag{26}$$

where γ_r, δ_s and ζ_t are a state, sector and year fixed effects respectively.[17] I do not include a plant fixed effect in the specification, since the theory predicts that capital growth ends once a plant reaches its optimal capital level. Hence, capital growth does not have a stable trend for a plant. I consider three different measures for $young_{irst}$, namely $[-\ln age_{irst}]$ and the indicator variables $1(age_{irst} \leq 5)$ and $1(age_{irst} < 10)$. The theoretical prediction is that the increase in competition due to dereservation leads to slower capital growth for young plants, i.e. $\beta_3 < 0$. Standard errors are clustered at the product level.

The estimation shows two main results (see Table 1). First, young plants have a higher average capital growth rate than older plants, and second, dereservation reduces the capital growth rate of the young plants, as predicted by the theory. These results are strongest and most significant for plants weakly younger than five years (columns 1-2). For instance in column 2, these plants have growth rates that are on average 0.024 log points higher ($p < 0.001$), but this growth rate falls by 0.022 log points after dereservation ($p = 0.001$). For plants younger than ten years, results are qualitatively similar, but the magnitudes are smaller.[18]

3.5 Full panel

In this subsection, I broaden the external validity of the estimation results by extending the analysis to the full panel of plants, instead of focusing only on incumbent plants whose products become dereserved. In this analysis, I set aside the model's predictions on markup misallocation and focus on capital growth, for two reasons. First, the prediction on capital growth is the model's most novel one, while the predictions on markup levels and dispersion have been ex-

[17] I do not include state-sector-year fixed effects in this specification since the number of state-sector-year groups is roughly one third of the number of observations, and therefore this type of fixed effects absorbs too much of the variation. However, in an additional specification I consider the full sample of plants, instead of only incumbent plants, and there I do include state-sector-year fixed effects γ_{rst}:

$$\begin{aligned} g(k_{irst}) = \gamma_{rst} + \beta_1 young_{irst} + \beta_2 incumb_{irst} + \beta_3 entrant_{irst} + \beta_4 incumb_{irst} * young_{irst} \\ + \beta_5 entrant_{irst} * young_{irst} + \beta_6 Deres_{irst-1} + \beta_7 Deres_{irst-1} * young_{irst} \\ + \beta_8 Deres_{irst-1} * entrant_{irst} + \beta_9 Deres_{irst-1} * entrant_{irst} * young_{irst} + \varepsilon_{irst}, \end{aligned} \tag{27}$$

The coefficient of interest is β_7, which estimates how the capital growth rate of incumbent firms changes after dereservation. Columns 2, 4 and 6 in Table 1 have the estimation output.

[18] The ASI data has incomplete product coverage for the years 1998 and 1999, and it changes the product classification after 2009. For these reasons, Appendix Table C.2 provides a robustness check for incumbent plants dereserved after 1999 and before 2010. Results are highly similar.

Table 1: Dereservation and capital growth for young plants

	Capital growth $g(k)_{it}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$Deres_{it-1} * 1(age_{it} \leq 5)$	-0.031***	-0.022***				
	(0.009)	(0.007)				
$Deres_{it-1} * 1(age_{it} < 10)$			-0.019***	-0.011*		
			(0.005)	(0.006)		
$Deres_{it} * [-\ln age_{it}]$					-0.008**	-0.002
					(0.003)	(0.003)
$1(age_{it} \leq 5)$	0.026***	0.024***				
	(0.008)	(0.003)				
$1(age_{it} < 10)$			0.008	0.013***		
			(0.005)	(0.003)		
$-\ln age_{it}$					0.001	0.004***
					(0.003)	(0.001)
$Deres_{it-1}$	0.000	-0.017***	-0.000	-0.020***	-0.022**	-0.019**
	(0.005)	(0.004)	(0.005)	(0.004)	(0.011)	(0.008)
State FE, Sector FE, Year FE	Yes	–	Yes	–	Yes	–
State-sector-year Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	163519	697183	163519	697183	161153	682237

The table estimates the impact of dereservation on capital growth of young plants, employing specifications (26) for columns 1, 3 and 5, and (27) for columns 2, 4 and 6. To check sensitivity to issues with the product classification (see Appendix Appendix B.2), Table C.2 provides a robustness check for incumbent plants dereserved after 1999 and before 2010. Standard errors, in parentheses, are clustered at the product level, which is the level at which dereservation status is defined. $^{*}p < 0.1;^{**} p < 0.05;^{***} p < 0.01$.

amined in previous research (e.g. Peters (2016); Schaumans and Verboven (2015)). Second, I will use the median markup in a market as my inverse measure of competition, and this choice does not allow me to examine markup misallocation. Importantly though, this competition measure is consistent with the model, since it predicts a monotonic relationship between the number of firms, which governs the degree of competition, and the first moments of the markup distribution.[19]

As my inverse measure of competition, I use the median markup at the region-sector-year level, $Median_{rst}[\ln \mu_{irst}]$, which is arguably exogenous from the plant's point of view. The markup is measured as

$$\mu_{irst} = \alpha_s^L \frac{S_{irst}}{w_{irst} L_{irst}}. \tag{28}$$

This markup measure is identical to equation (24), except that the elasticity α_s^L is now measured as a cost share at the sector level. Because the median markup will only enter in interaction terms in the specification below, $Median_{rst}[\ln \mu_{irst}]$ needs to be demeaned. To avoid results being driven by the measurement of α_s^L, demeaning happens within sectors. Hence, I am leveraging within-sector variation in the median markup, which is insensitive to α_s^L. Next, to ensure that

[19] From the point of view of the model, an alternative measure of competition could have been the number of firms. Note however that what matters for the degree of competition is not only the number of firms, but also market size, which depends on sectoral expenditure shares and income per capita, among others. The median markup incorporates these factors directly.

the median markup is plausibly exogenous to the individual plant, I restrict the sample to cases where at least seven plants are observed in a given region-sector-year. Finally, I normalize the median markup to standard deviation units. To examine its impact on young plants' capital growth in the full panel, I update the regression specification as follows:

$$g(k_{irst}) = \gamma_{rst} + \beta_1 young_{irst} + \beta_2 Median_{rst}[\ln \mu_{irst}] * young_{irst} + \varepsilon_{irst}, \tag{29}$$

where γ_{rst} is a state-sector-year fixed effect that, absorbs baseline variation in $Median_{rst}[\ln \mu_{irst}]$, among other shocks. The theoretical prediction is that less competition increases capital growth, i.e. $\beta_2 > 0$.

Heterogeneity along financial dependence So far the tests on capital growth have all implicitly assumed that the average plant in the sample is financially constrained. Now however, I turn to leveraging the empirical heterogeneity in the degree to which plants are financially constrained. Specifically, in sectors with higher levels of financial dependence, measured as $Fin\ Dep_s$, changes in the level of sector-level competition have a stronger impact on the rate of capital growth. I employ the standard Rajan and Zingales (1998) measure of sectoral financial dependence:

$$Fin\ Dep_s = \frac{Capital\ Expenditures_s - Cash\ Flow_s}{Capital\ Expenditures_s},$$

based on data for US sectors over the 1980's.[20] Here, $Fin\ Dep_s$ captures the share of external finance in a firm's investments in a setting with highly developed financial markets, namely the United States. The central idea in Rajan and Zingales (1998) is that in economies with less developed financial markets, such as India, financial constraints become especially binding in sectors with high levels of $Fin\ Dep_s$. In my setting, the model's prediction is that the impact of competition on capital growth for young plants is increasing with the degree of financial dependence ($\beta_4 > 0$):

$$\begin{aligned} g(k_{irst}) = &\gamma_{rst} + \beta_1 young_{irst} + \beta_2 Median_{rst}[\ln \mu_{irst}] * young_{irst} + \beta_3 young_{irst} * Fin\ Dep_s \\ &+ \beta_4 Median_{rst}[\ln \mu_{irst}] * young_{irst} * Fin\ Dep_s + \varepsilon_{irst}. \end{aligned} \tag{30}$$

Note that here again, γ_{rst} absorbs any common state-sector-year level variation, including in financial dependence and the median markup.

Estimation results The estimation results continue to be in line with the theoretical predictions (see Table 2). First, capital growth is higher for younger plants, with an average extra growth of 0.029 log points for plants weakly younger than five years ($p < 0.001$), and 0.013 log points for plants younger than ten years old ($p = .003$). Second, capital growth for young plants increases with the median markup. A one standard deviation increase in the median markup leads to an increase in capital growth by 0.009 ($p = 0.002$) and 0.006 log points ($p = 0.001$) for plants weakly

[20] I use the original Rajan and Zingales (1998) measures of financial dependence for ISIC Rev.2 sector definitions, except that I trim the financial dependence measure such that $Fin\ Dep_s \geq 0$. This ensures a clean identification of the effect of competition in the triple interaction term in specification (30). The ISIC Rev.2 sector definitions match closely with India's NIC 1987 sector definitions. The concordance between ISIC Rev.2 and NIC 1987 is provided by the Indian Statistical Office.

Table 2: Competition and capital growth of young plants

	Capital growth $g(k)_{irst}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$Median_{rst}[\ln \mu_{irst}] * 1(age_{irst} \leq 5)$	0.009***			0.006		
	(0.003)			(0.004)		
$Median_{rst}[\ln \mu_{irst}] * 1(age_{irst} < 10)$		0.006***			0.005*	
		(0.002)			(0.003)	
$Median_{rst}[\ln \mu_{irst}] * [-\ln age_{irst}]$			0.003***			0.001
			(0.001)			(0.002)
$Median_{rst}[\ln \mu_{irst}] * 1(age_{irst} \leq 5) * Fin\ Dep_s$				0.017**		
				(0.008)		
$Median_{rst}[\ln \mu_{irst}] * 1(age_{irst} < 10) * Fin\ Dep_s$					0.011**	
					(0.005)	
$Median_{rst}[\ln \mu_{irst}] * [-\ln age_{irst}] * Fin\ Dep_s$						0.009**
						(0.004)
$1(age_{irst} \leq 5)$	0.029***			0.019**		
	(0.005)			(0.008)		
$1(age_{irst} < 10)$		0.013***			0.003	
		(0.004)			(0.007)	
$-\ln age_{irst}$			0.002			-0.003
			(0.002)			(0.004)
$1(age_{irst} \leq 5) * Fin\ Dep_s$				0.031***		
				(0.012)		
$1(age_{irst} < 10) * Fin\ Dep_s$					0.029***	
					(0.010)	
$-\ln age_{irst} * Fin\ Dep_s$						0.016**
						(0.007)
State-sector-year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	620387	620387	607273	526675	526675	515139

The variable $Median_{rst}[\ln \mu_{irst}]$ is demeaned by sector, and then measured in standard deviation units. Columns 1-3 display estimation results for specification (29), while columns 4-6 show results for specification (30). To ensure that the median markup is plausibly exogenous to the individual plant, the sample is restricted to region-sector-years consisting of at least 7 plants. Standard errors, in parentheses, are clustered at the level of 3-digit sectors. $^{*}p < 0.1;^{**}p < 0.05;^{***}p < 0.01$.

younger than five and younger than ten years old, respectively. Both these patterns are amplified in sectors with high financial dependence: younger plants exhibit stronger capital growth and their capital growth is more strongly influenced by the median markup in these sectors. Both these findings are always highly statistically significant (see the triple interaction terms in columns 4-6). In terms of magnitude, moving from the tenth to the ninetieth percentile of financial dependence increases the measure by 0.9, and therefore a one-standard deviation increase in the median markup increases capital growth by 0.015 log points for plants younger than five years old.

3.6 Evidence on MRPK convergence

To provide further evidence that competition slows down convergence to unconstrained capital levels, in Appendix Section D, I show that the speed of convergence of MRPK (marginal revenue product of capital) is slower when competition is more intense. The advantage of focusing on MRPK convergence is that in contrast to capital growth of young plants, this analysis applies to plants of any age. To explain why even older firms may need to grow their capital, and how competition slows down this capital growth, I first write down a model with firm-level productivity shocks. After a positive productivity shock, firms decide to grow their capital, but the collateral constraint limits the speed at which they can do so. Less competition, and the associated higher markups, again facilitate faster capital growth.

Empirically, I show that MRPK convergence is slower for all three of the above empirical tests. It both slows down for incumbent plants after dereservation, and in settings where the median markup is lower. Moreover, the impact of the median markup is again larger in sectors where financial dependence is higher.

4 Quantification

After providing empirical support for the theoretical predictions, the final part of this paper sheds light on the quantitative importance of the positive and negative effects of competition on aggregate economic outcomes.

4.1 Simulation setup

In the quantification of the model, the parameter values are informed by a combination of typical values in the literature and data moments for the Indian manufacturing sector (see Table 3). First, the depreciation rate is set to a standard value of $\delta = 0.06$ (Buera, Kaboski, and Shin, 2011; Midrigan and Xu, 2014). Second, the elasticity of substitution is set to $\sigma = 4$ as in Bloom (2009) and Asker et al. (2014). This also ensures that the model's markup values will be fairly close to 1.34, which is the estimate for the median markup in the Indian manufacturing sector by De Loecker et al. (2016). Third, the discount factor has a standard value of $\beta = 0.95$, which ensures that the real interest rate is close to its median value of 5.47 in India. Fourth, the value for the output elasticity of capital (α) is based on the value for the labor share of revenue being $(1-\alpha)/\mu$ in the model, and empirically equal to $2/5$ in the Indian manufacturing sector (WIOD Socio-Economic Accounts). After following the above De Loecker et al. (2016) estimate by setting $\mu = 4/3$, this results in $\alpha = 7/15$.

In the simulations, the time period is a year, there are 100 sectors, and the number of firms in each sector (M) is always a multiple of 20. I drop the subscript in M_s from now on, since M remains constant across sectors in this section. Out of every 20 firms, each period one new firm is born and one firm dies. This closely mimics the birth rate in the Indian manufacturing sector, where on average 5.6% of plants are less than two years old. In steady state, the number of unconstrained firms and the number of constrained firms in each "age bin τ" is stable over time. To allow for a steady state even when M is low, in the simulations only firms older than a certain age T have a strictly positive probability of dying. By setting this age sufficiently high, I ensure that only unconstrained firms die, and that the share of constrained firms of each age $\tau \leq T$, as well as the share of firms with age $> T$, who are all unconstrained, are constant across steady states with different numbers of firms. [21] As described in the theory, all firms discount future consumption with a factor β. [22]

[21] In the simulations below, in the most extreme case, firms require 16 periods to become unconstrained. In that case, only firms of age 16 and beyond can die. In the baseline model however, firms always become unconstrained before age 10 and can be allowed to die earlier. In principle, it is possible to simulate economies where the death rate is constant over a firm's lifetime, as is assumed in the theoretical section. But then the requirement to have a stable number of firms in the constrained age bins necessitates a minimum number of firms that is substantially higher than 20. In such a case, the oligopolistic dynamics would become quantitatively negligible.

[22] As mentioned in the theory, I assume parameter values are such that $\beta = \eta\beta_f = 0.95$. With an aggregate survival rate of $\eta = 0.95$, this implies that $\beta_f = 1$. To simplify the exposition below, I am here making the simplifying assumption that firms do not take into account that the survival rate is age specific. To check the sensitivity of the results to this assumption, Appendix Figures E.5 and E.6 show simulation results when $\beta_f = 0.95$ and η_τ is age-specific. The basic

To calibrate λ (the collateral constraint) and ζ (the asset fraction at birth relative to unconstrained firms), I target an aggregate credit to revenue ratio and the size difference between newborn and mature firms. First, the credit to revenue ratio for medium, small, and micro manufacturing firms (MSME) is 0.155 in 2007, the earliest year with data available from the Indian Ministry of MSME. Second, the size difference between a plant at age 1 and at age 25 is 0.74 log points, a difference that remains relatively stable for older plants (see Figure C.7). By setting $\lambda = 0.23$ and $\zeta = 0.188$, the model matches these moments for $M = 40$.

Table 3: Parametrization

		Value	Reference
Parameters			
Depreciation rate	δ	0.06	Midrigan and Xu (2014)
Discount rate	β	0.95	Real interest rate
Elasticity of substitution	σ	4	Median markup
Capital elasticity	α	7/15	Labor share of revenue
Birth and death rate		1/20	Share of plants below age 2
Collateral constraint	λ	0.23	(Credit/Revenue) in MSME
Asset fraction at birth	ζ	0.188	Plants' size ratio
Number of firms	M_s	Multiples of 20	Simulation restriction
Calibration moments			
Real interest rate	$1 - 1/\beta$	5.47	World Bank (1990-2011)
Median markup	μ	1.34	De Loecker et al. (2016)
Labor share of revenue	$(1-\alpha)/\mu$	0.4	WIOD-SEA (1995-2009)
Share of plants with age < 2		5.6%	ASI (1990-2011)
(Credit/Revenue) in MSME		0.155	Indian Ministry of MSME (2007)
Plants' size ratio	$\ln\left(\frac{v_{\tau=25}}{v_{\tau=1}}\right)$	0.74	ASI (1990-2011)

The top panel displays the parameter values, and the bottom panel the moments in the data used to calibrate the parameters. Here, the real interest rate, the labor share of revenue, and the share of plants below age 2, are all median values over the indicated time period. For the credit to revenue ratio in Medium, Small and Micro Enterprises (MSME), 2007 is the earliest year with available data. Finally, the plants' size ratio is calculated after taking out year fixed effects.

The computer algorithm starts with finding the within-period general equilibrium, and subsequently updates firms' wealth levels. Before the next period, it randomly selects the firms who die and generates newborn firms. It then iterates over within-period equilibria until it converges on a stable distribution of capital and thereby obtains the steady state equilibrium. To find the within-period equilibrium, the algorithm first solves for the general equilibrium given demand

patterns in the simulations results persist in this robustness check: TFP falls with M, while aggregate capital, output and consumption increase.

elasticities ε_{ist} equal to σ. It then iteratively updates firms' market shares and demand elasticities, and solves for the associated general equilibrium until the distribution of market shares has converged.

4.2 Results for baseline model

Figure 2 visualizes what happens to markup levels and capital wedges when M increases. For the baseline value of the collateral constraint ($\lambda = 0.23$) and when there are 400 firms in each sector, all markup values – regardless of firm age – equal 1.34, up to a rounding error (see Panel b). Here, market shares are tiny, and as a result markups are only just above $4/3$, their level under monopolistic competition. In contrast, when there are merely 20 firms in each sector, the older and larger firms charge a markup of 1.410, whereas the newborn firms set their markup at 1.369. Hence, markup levels and markup dispersion are higher when there are fewer firms in each sector (see Proposition 1). Note here also that going from 20 to 40 firms already closes roughly half the gap with the monopolistically competitive markups. Hence, the marginal impact of competition is largest when the baseline number of firms is lower.

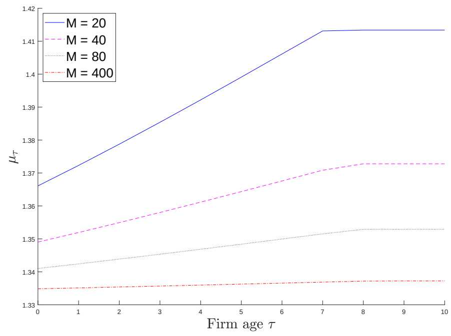
The markup distributions are similar when firms have no access to finance ($\lambda = 0$, Panel a), except on two dimensions. First, since firms will take longer to grow their capital to the unconstrained level, the firms who set the highest markup are older and fewer in number. Second, the range between the minimum and maximum markup, a measure of markup dispersion, is slightly larger. For instance, when $M = 20$, the difference in markups is 4.73 percentage points when $\lambda = 0$, versus 4.08 percentage points when $\lambda = 0.23$. This is because the unconstrained firms have a larger market share since the newborn firms have lower capital levels and it takes them longer to grow out of their constraint.

Panels (c) and (d) show how capital growth slows down with competition. Specifically, they visualize how fast firms close their "capital wedge," measured as k_τ/k_{10}, which is the ratio of the capital levels of firms of age τ and age 10 respectively, where the latter firm is unconstrained. When $\lambda = 0.23$ (Panel d), the effect of competition is modest but non-negligible: when $M = 20$, capital grows by 288% in the first 5 periods, and this growth drops to 257% and 233% for $M = 40$ and $M = 400$ respectively. While the cumulative capital growth rate drops a substantial 55 percentage points when M increases from 20 to 400, firms only take one year longer to become unconstrained in the latter case, namely 7 instead of 6 years. This is because the annual capital growth rate remains elevated. Not surprisingly, these effects are somewhat amplified when firms have no access to external finance (Panel c), which implies that capital at age 0 is restriced to firms' initial wealth and that retained earnings are the only source of capital growth. At age 7, firms in all equilibria are still constrained, and the cumulative capital growth up to that age is 417%, 370% and 334% for M equal to 20, 40 and 400 respectively. Here, the marginal impact of competition remains largest for low baseline levels of M, since going from 20 to 40 firms bridges more than half the gap in capital growth rates between equilibria with 20 versus 400 firms.

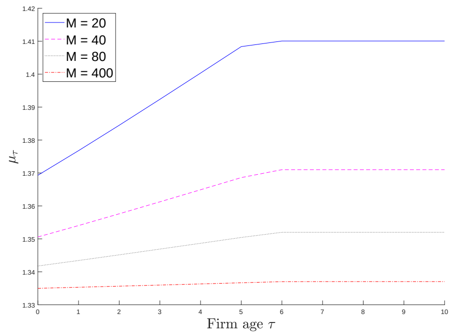
Figure 3 examines how the combination of lower markup levels, markup dispersion and capital growth rates is reflected in aggregate economic variables. First, TFP always declines with competition when access to external finance is imperfect (Panel a; $\lambda < 1$). Hence, for allocative efficiency of inputs, the negative effect of competition on capital growth dominates the positive effect that arises from reduced markup dispersion. Importantly though, the negative effect on

Figure 2: Impact of competition on capital wedges and markups

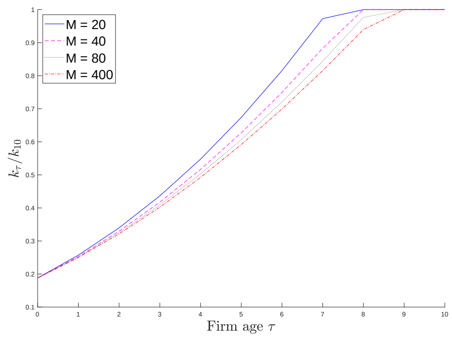
(a) Markup levels when $\lambda = 0$



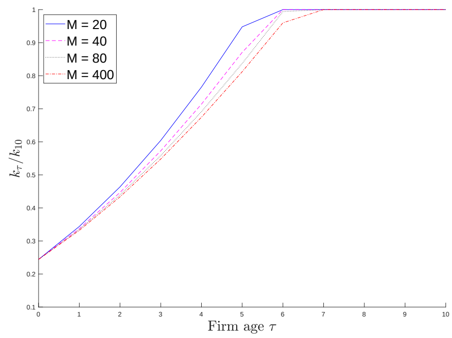
(b) Markup levels when $\lambda = 0.23$



(c) Capital wedges when $\lambda = 0$



(d) Capital wedges when $\lambda = 0.23$



The figure displays the simulation results for an increase in the number of firms M in all sectors. Panels (a) and (b) plot the distribution of markups μ_τ for firms of age τ for $\lambda = 0$ and $\lambda = 0.23$ respectively, while Panels (c) and (d) plot the distribution of capital wedges for the same respective λ values. A capital wedge is measured as the ratio of the capital level of firms of age τ and age 10, where the latter firm is always unconstrained in the current simulations. Since here all firms above age 10 are unconstrained, these firms are omitted from the figure. The values for λ are two benchmark values: $\lambda = 0$ entails no access to external finance, while $\lambda = 0.23$ is the preferred, calibrated value.

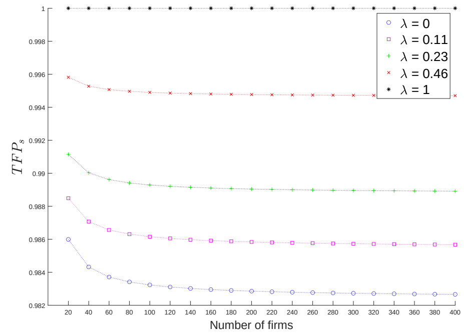
allocative efficiency tends to be small. Going from 20 to 400 firms in a sector, the maximal decline in TFP is 0.33 percentage points (when $\lambda = 0$), and for the baseline value of $\lambda = 0.23$, the decline is 0.22 percentage points. As explained in Section 2.3, there is always perfect allocative efficiency when $\lambda = 1$, regardless of M.[23]

While TFP declines, aggregate capital, output and consumption all rise with M. The largest gains are evident for aggregate capital, which increases by 8.5% when $\lambda = 0.23$. This decrease is driven by the decline in markups, which leads firms to accumulate more capital. Since the marginal impact of competition is largest when the baseline number of firms is low, the marginal increase in aggregate capital is also highest when M is low.

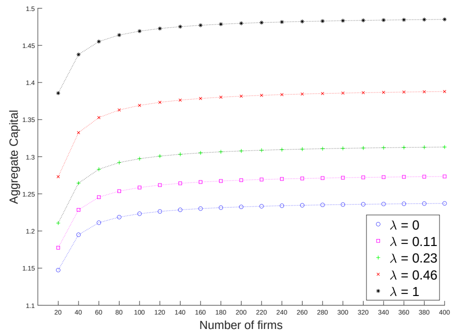
[23] It is also clear that TFP is monotonically increasing in λ, a finding reflecting the pattern in Midrigan and Xu (2014). In their benchmark case, the decline in TFP due to financing constraints is just 0.3%, whereas here it is ca. 1%. One reason why I obtain a larger loss, is that the calibrated value for the collateral constraint is lower in my case ($\lambda = 0.23$ versus $\lambda = 0.86$). When the collateral constraint declines in Midrigan and Xu (2014), the TFP loss in their model eventually becomes larger than in this paper. This is partly due to their model generating a much larger fraction of constrained firms when $\lambda = 0$, namely 83% versus 45% in this paper. The fact that mature firms are not exposed to positive shocks in the current model helps explain this difference.

Figure 3: Impact of competition on aggregate outcomes

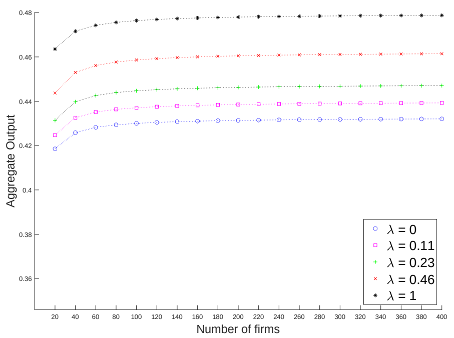
(a) Allocative efficiency



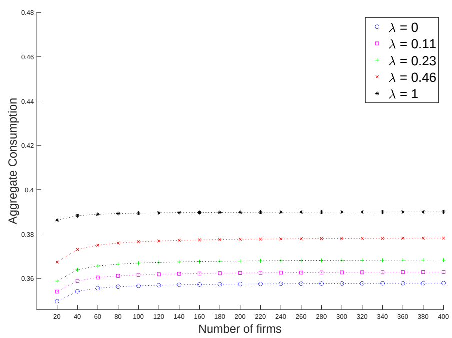
(b) Aggregate Capital



(c) Output



(d) Consumption



The figure displays the simulation results for an increase in the number of firms M in all sectors, for different values of λ. The preferred, calibrated value for λ is 0.23, and the values 0.11 and 0.46 are (roughly) half and twice that value, respectively. No access and perfect access to finance respectively imply $\lambda = 0$ and $\lambda = 1$. Panel (a) shows the impact of M on TFP_s; Panel (b) on aggregate capital across all firms in all sectors; Panel (c) on output of the final good Q_F; and Panel (d) on the sum of aggregate worker and firm consumption.

Importantly, this sizable increase in capital more than offsets the negative impact of the small decline in TFP, since both aggregate output and consumption monotonically increase with competition. Specifically, for $\lambda = 0.23$, going from $M = 20$ to $M = 400$ increases output by 3.63%, and aggregate consumption, which sums all worker and firm consumption, by 2.65%. The smaller relative increase in output versus capital reflects the capital share in production, while the uptick in aggregate consumption lies below the increase in output because the higher capital level requires higher aggregate investment to replace depreciated capital.

Interestingly, among the displayed values for λ, the increase in aggregate consumption when M rises from 20 to 400, is largest for $\lambda = 0.46$, and smallest for $\lambda = 1$ (2.95% vs. 0.97%). This is a result of two opposing forces: the decline in TFP is largest when λ is low, while at the same time, decreasing returns to capital imply that the marginal value of aggregate capital is highest in this case. These two opposing forces lead to an inverted U-shape in the consumption gain from increased competition as a function of λ.

Comparative statics on σ and α How does the value for the elasticity of substitution (σ), affect the quantitative analysis? Since σ determines the lower bound on the markup, it matters substantially for both the markup distribution and the speed of capital growth. For instance, when $\sigma = 3$, markup values are between 1.5 and 1.6 (Appendix Figure E.1, Panel a), and firms become unconstrained by age 5 at the latest (Panel c). In contrast, when $\sigma = 5$, markups are ca. 15% lower (Panel b) and as a consequence firms take 2-3 periods longer to grow out of their capital constraint (Panel d). The lower baseline level of markups also implies that the impact of increased competition is larger. When $\sigma = 3$, the maximal difference in the cumulative capital growth rate up until age τ is 42 percentage points (at age 4), whereas it is 70 percentage points when $\sigma = 5$ (at age 6). This pattern is reflected in a somewhat stronger negative impact of competition on TFP when σ is higher (see Appendix Figure E.2, Panels a and b). In addition, a lower σ is associated with a slightly higher increase in aggregate capital as M increases, which is due to larger drop in markups (see again Figure E.1, Panels a and b). The smaller drop in TFP and the larger increase in aggregate capital for low σ translate into a slightly larger consumption gain from increased M, namely 2.93% when $\sigma = 5$ versus 2.39% when $\sigma = 3$, each time for $\lambda = 0.23$.

The value for the capital share α affects how long it takes firms to grow out of their capital constraint: a higher α increases the unconstrained capital level (see equation (15)) and implies a steeper marginal cost curve for constrained firms (see equation (18)). For the low value $\alpha = 4/15$, all firms are unconstrained within 4 periods, while for a high value of $\alpha = 10/15$, firms require 7, 8 or 9 periods to become unconstrained, depending on the degree of competition (see Appendix Figure E.3, Panels c and d). The capital share also affects the magnitude of the capital wedges and their importance in TFP. For $\alpha = 4/15$, firms quickly become unconstrained. As a result, the capital wedges are relatively small, which translates into a relatively high TFP (see Appendix Figure E.4, Panel a). In contrast, when $\alpha = 10/15$, capital wedges are larger, and widen further with competition, which translates into a lower TFP that declines more strongly with M (Panel b).

Both for high and low values of α, aggregate capital increases with M due to the decline in markup levels (Panels c and d), and this translates into increased output (Panels e and f), offsetting the decline in TFP. Typically, these trends are associated with higher consumption (Panels g and h). However, there is an exception: when α is low and access to finance is perfect, the increased aggregate investment rate associated with higher M becomes suboptimally high, as the marginal cost of replacing depreciated capital outweighs the marginal increase in output.

4.3 Fixed operating cost

Model setup Thus far, I have been abstracting from two common components of models on firm dynamics: love of variety on the demand side, and a fixed cost of operation on the supply side. As is well known, the combination of these components introduces a trade-off between increased variety and reduced scale economies when increasing the number of firms (Dixit and Stiglitz, 1977). I now examine how the interaction of these forces with financial constraints affects the benefits of competition.

I incorporate love of variety by dropping the $M_s^{1/(1-\sigma)}$ term in equation (2) and updating the demand, revenue, and price functions accordingly. Next, I introduce a fixed operating cost of F labor units by updating equation (14) for a firm owner's net real wealth as follows

$$a_{ist} \equiv \pi_{ist}(l_{ist}, k_{ist}) + (1-\delta)k_{ist} - (1+r^d)d_{ist} - w_t F.$$

Given the fixed cost, firms make a choice to operate or not. At the start of a period, if $(1 + r^d)a_{ist-1} > \pi_{ist}(l_{ist}, k_{ist}) + (1-\delta)k_{ist} - w_t F$, then they decide not to operate and to lend out their wealth to others instead. In my simulations however, in steady state all firms always choose to operate. I set the value for the new parameter at $F = 0.0235$. This is based on Midrigan and Xu (2014)'s calibration of the share of the labor force employed to cover the fixed cost (9.4%), applied to a version of this model with 40 firms in each of the 100 sectors.

Results In this extended model, an increase in M has drastic effects on capital growth (see Figure 4, Panels c and d). For the baseline value of $\lambda = 0.23$, it takes firms 6 periods to become unconstrained when $M = 20$, but this increases to 14 periods when $M = 140$. For $\lambda = 0$ the capital growth at $M = 20$ is slower, and the increase in competition leads to a similar slowdown of an additional 8 periods. This drop in the capital growth rate is particularly substantial when recalling that in the baseline model, the slowdown induced by competition was at most one period.

The amplified slowdown in capital growth is not driven by changes in the markup distribution, since this distribution is mostly unaltered in comparison to the baseline model (see Panels a and b).[24] Instead, the slowdown in capital growth is driven by the reduction of economies of scale. As M increases, firms of all ages are smaller, and as a result, the fixed cost share of revenue (wF/v_τ) rises (Panels e and f). As a consequence, retained earnings fall sharply, which results in drastically slower capital growth.

This decline in capital growth has substantial consequences on aggregate outcomes (see Figure 5). First, allocative efficiency declines steadily with M: by 1.12 percentage points for $\lambda = 0.23$, going from 20 to 140 firms. Second, aggregate capital, which was monotonically increasing in M in the baseline model, is now an inverted U-shaped function of M. At first it increases with M due to the decline in markups and love of variety effects on the final good. However, eventually the severe slowdown in firm-level capital growth drags down aggregate capital accumulation. Third, aggregate output and consumption follow a similar pattern. Variety gains and lower markups lead to an initial increase in output and consumption, but reduced scale economies and their financial impact on capital growth eventually decrease these aggregate outcomes. The slowdown in capital growth matters here, since peak consumption is reached for a strictly lower M when $\lambda < 1$. When $\lambda = 0$, optimal M is 120 number of firms, while for $\lambda = 0.11, 0.23$ and 0.46 it is 140. For $\lambda = 1$, optimal M is still strictly higher.

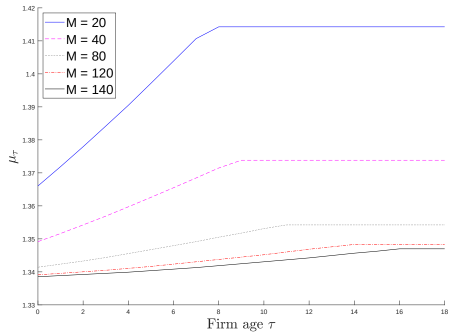
5 Conclusion

This paper started with the motivating example of India, a country that implemented multiple liberalization reforms, but where misallocation of resources remained persistent. I explained how

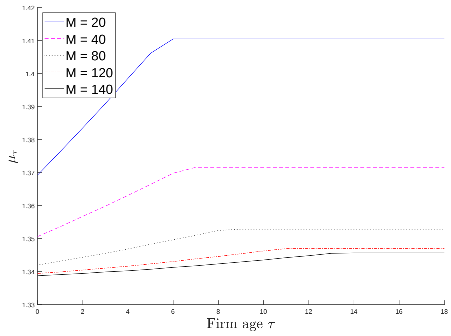
[24]Note that $M = 140$ is the maximum number of firms for which the economy has a steady state when $\lambda = 0$, since for $M > 140$, firms take more than 18 periods to become unconstrained. As a result, the number of unconstrained firms as a share of 20 (the minimum number of firms) is unstable over time. While it is possible to simulate a steady state for $M > 140$ by increasing the minimum number of firms (e.g. work with multiples of 30 firms for $M = 150$), the addition of extra age bins would complicate the comparison with the baseline model.

Figure 4: Impact of competition on markups and capital growth

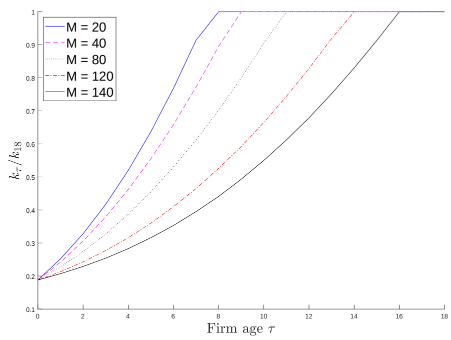
(a) Markup levels when $\lambda = 0$



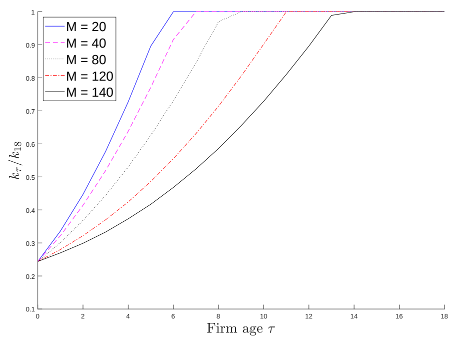
(b) Markup levels when $\lambda = 0.23$



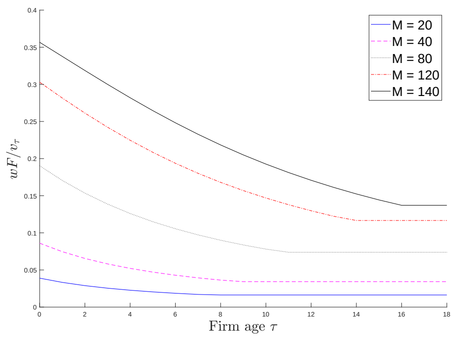
(c) Capital wedges when $\lambda = 0$



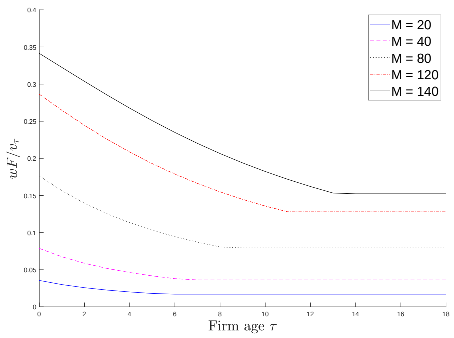
(d) Capital wedges when $\lambda = 0.23$



(e) Fixed cost share of revenue when $\lambda = 0$



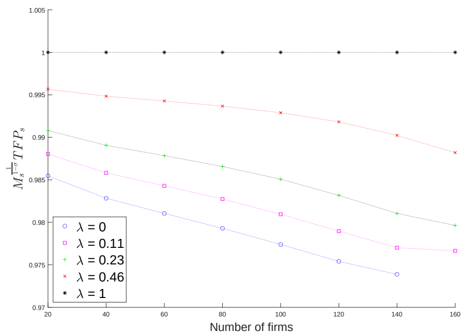
(f) Fixed cost share of revenue when $\lambda = 0.23$



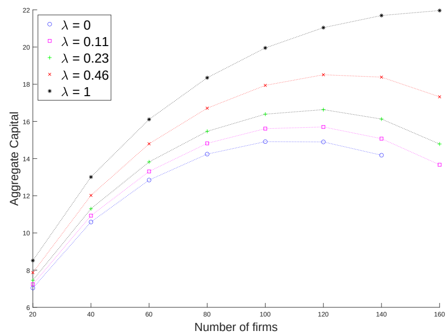
The figure displays the simulation results for an increase in the number of firms M in all sectors for the model with love of variety and fixed operating costs. Panels (a) and (b) plot the distribution of markups μ_τ for firms of age τ for $\lambda = 0$ and $\lambda = 0.23$ respectively, Panels (c) and (d) plot the distribution of capital wedges, and Panels (e) and (f) plot the share of the fixed cost in revenue (wF/v_τ), each time for the same respective λ values. A capital wedge is measured as the ratio of the capital level of firms of age τ and age 18, where the latter firm is always unconstrained in the current simulations. Since here all firms above age 18 are unconstrained, these firms are omitted from the figure. The values for λ are two benchmark values: $\lambda = 0$ entails no access to external finance, while $\lambda = 0.23$ is the preferred, calibrated value.

Figure 5: Impact of competition on aggregate variables

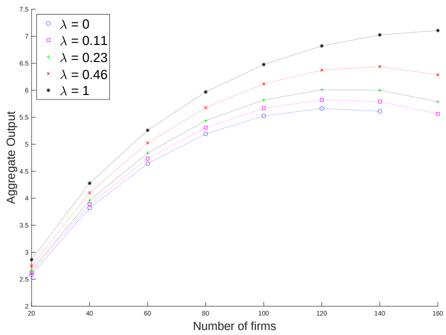
(a) Allocative Efficiency



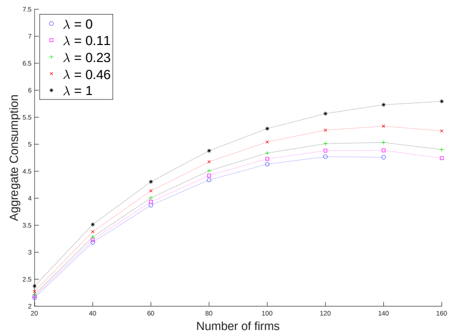
(b) Aggregate Capital



(c) Output



(d) Consumption



The figure displays the simulation results for an increase in the number of firms M in all sectors, for different values of λ. The preferred, calibrated value for λ is 0.23, and the values 0.11 and 0.46 are (roughly) half and twice that value, respectively. No access and perfect access to finance respectively imply $\lambda = 0$ and $\lambda = 1$. Panel (a) shows the impact of M on $M^{1/(1-\sigma)}TFP_s$, where the term $M^{1/(1-\sigma)}$ takes out the love of variety gain to focus on allocative efficiency. Panel (b) displays the impact of M on aggregate capital across all firms in all sectors; Panel (c) on output of the final good Q_F; and Panel (d) on the sum of aggregate worker and firm consumption. The maximum number of firms in the figure is 160, since when $\lambda = 0.46$, firms take longer than 18 periods to become unconstrained when $M > 160$. As a consequence, the number of firms in age bin τ of constrained firms is never stable over time such that a steady state is impossible. For $\lambda = 0$, a steady state is impossible for $M > 140$. It is possible to simulate a steady state by increasing the minimum number of firms beyond 20 to allow for more age bins. I do not perform this exercise since I choose to work with multiples of 20 throughout the quantitative analysis.

increased competition may not have the anticipated effect of improving allocative efficiency, since it slows down capital growth of financially constrained firms and thereby amplifies dispersion in marginal products of capital. When firms have fixed costs of operating, the slowdown in capital growth eventually becomes sufficiently large to offset the marginal benefits of increasing competition further. As a result, reduced access to finance leads to a decline in the optimal degree of competition in the economy.

These findings are particularly salient for countries where firms' access to finance is highly limited. There, the benefits from pro-competitive reforms may be underwhelming. While this insight echoes the conclusion of the old infant-industry argument, the policy implications are starkly different. After all, this paper shows that there are always gains from competition, as long as access to finance is sufficiently broad. For practicioners, the take-away could therefore be to first optimize financial access for firms, before enhancing competition in the real economy.

References

Aghion, P., U. Akcigit, and P. Howitt (2014). What do we learn from Schumpeterian growth theory? In *Handbook of Economic Growth*, Volume 2, pp. 515–563. Elsevier.

Aghion, P., R. Burgess, S. J. Redding, and F. Zilibotti (2008). The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India. *American Economic Review 98*(4), 1397–1412.

Akcigit, U., H. Alp, and M. Peters (2016). Lack of selection and limits to delegation: Firm dynamics in developing countries. *NBER working paper*.

Alfaro, L. and A. Chari (2014). Deregulation, Misallocation, and Size: Evidence from India. *The Journal of Law and Economics 57*(4), 897–936.

Allcott, H., A. Collard-Wexler, and S. D. O'Connell (2016). How do electricity shortages affect industry? Evidence from India. *American Economic Review 106*(3), 587–624.

Amiti, M., O. Itskhoki, and J. Konings (2016). International shocks and domestic prices: how large are strategic complementarities? *NBER 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.

Asturias, J., M. García-Santana, and R. Ramos (2019). Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india. *Journal of the European Economic Association 17*(6), 1881–1940.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *The American Economic Review 98*(5), 1998–2031.

Balasundharam, V. (2018). Bottlenecks versus ripple effects: The role of linkages in india's product market liberalization. *working paper*.

Banerjee, A., S. Cole, and E. Duflo (2005). Bank Financing in India. In *India's and China's recent experience with reform and growth*, pp. 138–157. Springer.

Banerjee, A. V. and E. Duflo (2014). Do firms want to borrow more? Testing credit constraints using a directed lending program. *The Review of Economic Studies 81*(2), 572–607.

Bils, M., P. J. Klenow, and C. Ruane (2020). Misallocation or mismeasurement? Technical report, National Bureau of Economic Research.

Blanchard, O. J. and N. Kiyotaki (1987). Monopolistic competition and the effects of aggregate demand. *The American Economic Review*, 647–666.

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica 77*(3), 623–685.

Boehm, J., S. Dhingra, and J. Morrow (2019). The comparative advantage of firms.

Boehm, J. and E. Oberfield (2018). Misallocation in the market for inputs: Enforcement and the organization of production. *NBER working paper*.

Bollard, A., P. J. Klenow, and G. Sharma (2013). India's Mysterious Manufacturing Miracle. *Review of Economic Dynamics 16*(1), 59–85.

Buera, F. J., J. P. Kaboski, and Y. Shin (2011). Finance and development: A tale of two sectors. *American economic review 101*(5), 1964–2002.

Buera, F. J., J. P. Kaboski, and Y. Shin (2015). Entrepreneurship and financial frictions: A macrodevelopment perspective. *Annual Review of Economics 7*(1), 409–436.

Caggese, A. and A. Pérez-Orive (2017). Capital misallocation and secular stagnation. *working paper*.

Carreira, C. and F. Silva (2010). No deep pockets: Some stylized empirical results on firms'financial constraints. *Journal of Economic Surveys 24*(4), 731–753.

Chari, A. (2011). Identifying the aggregate productivity effects of entry and size restrictions: An empirical analysis of license reform in India. *American Economic Journal: Economic Policy 3*(2), 66–96.

De Loecker, J. and J. Eeckhout (2018). Global market power. Technical report, National Bureau of Economic Research.

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. *The American Economic Review 102*(6), 2437–2471.

Dhingra, S. and J. Morrow (2019). Monopolistic competition and optimum product diversity under firm heterogeneity. *Journal of Political Economy 127*(1), 000–000.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *The American economic review 67*(3), 297–308.

Epifani, P. and G. Gancia (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics 83*(1), 1–13.

Foellmi, R. and M. Oechslin (2016). Harmful pro-competitive effects of trade in presence of credit market frictions. *manuscript, University of St. Gallen*.

Fort, T. C., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2013). How firms respond to business cycles: The role of firm age and firm size. *IMF Economic Review 61*(3), 520–559.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies 38*(1), 1–12.

García-Santana, M. and J. Pijoan-Mas (2014). The reservation laws in India and the misallocation of production factors. *Journal of Monetary Economics 66*, 193–209.

Gilbert, R. (2006). Looking for mr. schumpeter: Where are we in the competition-innovation debate? In *Innovation Policy and the Economy, Volume 6*, pp. 159–215. The MIT Press.

Guner, N., G. Ventura, and Y. Xu (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics 11*(4), 721–744.

Hall, R. (1986). Market structure and macroeconomic fluctuations. *Brookings papers on economic activity 1986*(2), 285–338.

Harrison, A. and A. Rodríguez-Clare (2010). Trade, foreign investment, and industrial policy for developing countries. In *Handbook of development economics*, Volume 5, pp. 4039–4214. Elsevier.

Hottman, C. J., S. J. Redding, and D. E. Weinstein (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics 131*(3), 1291–1364.

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

Itskhoki, O. and B. Moll (2019). Optimal development policies with financial frictions. *Econometrica 87*(1), 139–173.

Jungherr, J. and D. Strauss (2017). A Blessing in Disguise? Market Power and Growth with Financial Frictions. *working paper*.

Kehrig, M. and N. Vincent (2017). Do firms mitigate or magnify capital misallocation? evidence from plant-level data. *working paper*.

Levine, R. (2005). Finance and growth: theory and evidence. *Handbook of economic growth 1*, 865–934.

Mankiw, N. G. and M. D. Whinston (1986). Free entry and social inefficiency. *The RAND Journal of Economics*, 48–58.

Martin, L. A., S. Nataraj, and A. E. Harrison (2014). In with the Big, Out with the Small: Removing Small-Scale Reservations in India. *NBER Working Paper Series*.

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.

Midrigan, V. and D. Y. Xu (2014). Finance and Misallocation: Evidence from Plant-Level Data. *American Economic Review 104*(2), 422–458.

Mohan, R. (2002). Small-scale industry policy in india. *Economic policy reforms and the Indian economy 213*.

Moll, B. (2014). Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *The American Economic Review 104*(10), 3186–3221.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 1417–1426.

Peters, M. (2016). Heterogeneous Mark-ups, Growth and Endogenous Misallocation. *mimeo, Yale University*.

Rajan, R. and L. Zingales (1998). Financial dependence and growth. *The American Economic Review 88*(3), 559–586.

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

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and statistics 71*(1), 135–158.

Schaumans, C. and F. Verboven (2015). Entry and competition in differentiated products markets. *Review of Economics and Statistics 97*(1), 195–209.

Sivadasan, J. (2009). Barriers to competition and productivity: Evidence from India. *The BE Journal of Economic Analysis & Policy 9*(1).

Tewari, I. and J. Wilde (2017). Product Scope and Productivity: Evidence from India's Product Reservation Policy. *working paper, University of South Florida*.

Appendix A Proofs for Lemma 2 and Proposition 1

A.1 Proof for Lemma 2

As a preliminary, note that the expression for capital growth $G_{s\tau}$ in Lemma 1 can be written as

$$G_{s\tau} \equiv \frac{a_{s\tau}}{a_{s-1}} = \frac{\pi_{s\tau}}{a_{s-1}} + \frac{1-\delta-\lambda-r^d}{1-\lambda} G_{s\tau-1} \tag{31}$$

Given this, I will proceed by induction. In Step 1, I demonstrate that

$$(m_s^u > m_s^{u'}) \implies ((m_{s0} > m'_{s0}) \wedge (G_{s0} > G'_{s0})) \, ,$$

and afterwards I demonstrate the inductive step that

$$(m_s^u > m_s^{u'}) \wedge (G_{s\tau-1} > G'_{s\tau-1}) \implies ((m_{s\tau} > m'_{s\tau}) \wedge (G_{s\tau} > G'_{s\tau})) \, .$$

The first step and the inductive step together imply that Lemma 2 holds.

Step 1.A I first demonstrate that $(m_s^u > m_s^{u'}) \implies (m_{s0} > m'_{s0})$, employing a proof by contradiction. Therefore, suppose to the contrary that $(m_s^u > m_s^{u'}) \wedge (m_{s0} \leq m'_{s0})$.[25] First defining the output gap:

$$\mathcal{G}_{s0} \equiv \left(\frac{y_s^u}{y_{s0}}\right)^{\frac{\sigma-1}{\sigma}},$$

which from equation (7) implies that $m_{s0} = m_s^u/\mathcal{G}_{s0}$. Hence the supposition implies that the output ratio is higher in the former equilibrium:

$$(m_s^u > m_s^{u'}) \wedge (\mathcal{G}_{s0} > \mathcal{G}'_{s0}).$$

Combining the expression for the price $p_{ist} = MC_{ist}\mu_{ist}$, and optimal output being a downward sloping function of a firm's price, as implied by equation (3), we have that

$$\mathcal{G}_{s0} = \left(\frac{MC_{s0}\mu_{s0}}{MC_s^u \mu_s^u}\right)^{\sigma-1}.$$

Hence, with $\sigma > 1$,

$$(\mathcal{G}_{s0} > \mathcal{G}'_{s0}) \implies \left(\left(\frac{MC_{s0}}{MC_s^u} > \frac{MC'_{s0}}{MC_s^{u'}}\right) \vee \left(\frac{\mu_{s0}}{\mu_s^u} > \frac{\mu'_{s0}}{\mu_s^{u'}}\right)\right).$$

In point (i), note that $MC_s^u = MC_s^{u'}$, and $MC_{s0} = \frac{w}{1-\alpha}(l_{s0}/k_{s0})^\alpha$, $\left(\frac{MC_{s0}}{MC_s^u} > \frac{MC'_{s0}}{MC_s^{u'}}\right) \implies \left(\frac{l_{s0}}{k_{s0}} > \frac{l'_{s0}}{k'_{s0}}\right)$. Combined with $k_{s0} = \zeta_s k_s^u$, this entails that $(\mathcal{G}_{s0} > \mathcal{G}'_{s0}) \wedge \left(\frac{MC_{s0}}{MC_s^u} > \frac{MC'_{s0}}{MC_s^{u'}}\right)$ implies a contradiction. In point (ii), since $(m_s^u > m_s^{u'}) \implies \frac{\mu_s^u}{\mu_s^{u'}} > 1$, $\left(\frac{\mu_{s0}}{\mu_s^u} > \frac{\mu'_{s0}}{\mu_s^{u'}}\right)$ implies that $\mu_{s0} > \mu'_{s0}$, which is a contradiction with $(m_s^u > m_s^{u'}) \wedge (\mathcal{G}_{s0} > \mathcal{G}'_{s0})$.

The combination of point (i) and (ii) entails a contradiction with the supposition. Hence, its opposite must be true, which proves that $(m_s^u > m_s^{u'}) \implies (m_{s0} > m'_{s0})$.

Step 1.B In this step 1.B, I show that $(m_{s0} > m'_{s0}) \implies (G_{s0} > G'_{s0})$. Given equation (31), this is the case if

$$(m_{s0} > m'_{s0}) \implies \left(\frac{\pi_{s0}}{a_{s-1}} > \frac{\pi'_{s0}}{a'_{s-1}}\right).$$

Here, note that for a newborn constrained firm we have:[26]

$$\frac{\pi_{s0}}{a_{s-1}} = \frac{\pi_{s0}}{k_{s0}} = (\mu_{s0} - (1-\alpha))\frac{l_{s0}}{k_{s0}}\frac{w}{1-\alpha}, \tag{32}$$

---

[25] Recall that $\neg(p \implies q) \iff (p \wedge \neg q)$.

[26] Revenue net of labor cost is $\pi_{s\tau} = (p_{s\tau} - ALC_{s\tau})\, y_{s\tau}$, where $ALC_{s\tau}$ is the average cost of labor input. It is useful to rewrite this as:

$$\pi_{s\tau} = \left(\mu_{s\tau} - \frac{ALC_{s\tau}}{MC_{s\tau}}\right) y_{s\tau} MC_{s\tau},$$

Given the Cobb-Douglas production function, we have that total labor costs for any given quantity $\bar{y}_{s\tau}$ are $TLC(\bar{y}_{s\tau}) = wl(\bar{y}_{s\tau})$. For constrained firms, setting $\bar{y}_{s\tau}$ directly implies setting the amount of labor in the following function: $l(\bar{y}_{s\tau}) = \left(\frac{\bar{y}_{s\tau}}{k_{s\tau}^\alpha}\right)^{\frac{1}{1-\alpha}}$, such that $TLC(\bar{y}_{s\tau}) = w\left(\frac{\bar{y}_{s\tau}}{k_{s\tau}^\alpha}\right)^{\frac{1}{1-\alpha}}$. Hence, $ALC_{s\tau}(\bar{y}_{s\tau}) = w\left(\frac{\bar{y}_{s\tau}^\alpha}{k_{s\tau}^\alpha}\right)^{\frac{1}{1-\alpha}}$, and given these firms' marginal cost is as in equation (18), $\frac{ALC_{s\tau}}{MC_{s\tau}} = (1-\alpha)$. Together with $\frac{\bar{y}_{s\tau}}{k_{s\tau}^\alpha} = l_{s\tau}^{1-\alpha}$, this implies that:

$$\pi_{s\tau} = (\mu_{s\tau} - (1-\alpha))\, l_{s\tau}\frac{w}{1-\alpha}.$$

Then, consider three cases; first $\frac{l_{s0}}{k_{s0}} \geq \frac{l'_{s0}}{k'_{s0}}$, and second $\frac{l_{s0}}{k_{s0}} < \frac{l'_{s0}}{k'_{s0}}$.

- Case (i): suppose $\frac{l_{s0}}{k_{s0}} \geq \frac{l'_{s0}}{k'_{s0}}$. Since $(m_{s0} > m'_{s0}) \iff (\mu_{s0} > \mu'_{s0})$, this case immediately implies that $(\frac{\pi_{s0}}{a_{s-1}} > \frac{\pi'_{s0}}{a'_{s-1}})$.
- Case (ii): suppose the firm, following the reaction function in Equation (19), chooses optimally to have $\frac{l_{s0}}{k_{s0}} < \frac{l'_{s0}}{k'_{s0}}$. Since the reaction function in Equation (19) maximizes the revenue net of labor costs of the firm at age 0, and since $\frac{l_{s0}}{k_{s0}} \geq \frac{l'_{s0}}{k'_{s0}}$ is within the firm's choice set, setting $\frac{l_{s0}}{k_{s0}} < \frac{l'_{s0}}{k'_{s0}}$ implies that its revenue net of labor costs is weakly higher than under case (i), so $(\frac{\pi_{s0}}{a_{s-1}} > \frac{\pi'_{s0}}{a'_{s-1}})$ continues to hold. What happens in this case (ii), is that the firm optimally restricts output to maximize revenue, and this restricted output is optimal given that its demand is more inelastic.[27]

Inductive Step A I now demonstrate that the inductive step holds. I start with showing

$$(m_s^u > m_s^{u'}) \wedge (G_{s\tau-1} > G'_{s\tau-1}) \implies (m_{s\tau} > m'_{s\tau}).$$

The proof procedes analogous as in Step 1.A. First, suppose to the contrary that $(m_s^u > m_s^{u'}) \wedge (G_{s\tau-1} > G'_{s\tau-1}) \wedge (m_{s\tau} \leq m'_{s\tau})$. Given the definition of the output ratio, we have that $m_s^u = \mathcal{G}_{s\tau} y_{s\tau}$. Hence the supposition implies that the output ratio falls:

$$(m_s^u > m_s^{u'}) \wedge (G_{s\tau-1} > G'_{s\tau-1}) \wedge (\mathcal{G}_{s\tau} \leq \mathcal{G}'_{s\tau}).$$

We again have that $\mathcal{G}_{s\tau} = (MC_{s\tau}\mu_{s\tau}/MC_s^u\mu_s^u)^{\sigma-1}$, such that either relative marginal costs or relative markups need to be higher in the former equilibrium. This again results in a contradiction.

In point (i), note that because $MC_s^u = MC_s^{u'}$, and $MC_{s\tau} = \frac{w}{1-\alpha}(l_{s\tau}/G_{s\tau}k_{s0})^{\alpha}$, $\left(\frac{MC_{s\tau}}{MC_s^u} > \frac{MC'_{s\tau}}{MC_s^{u'}}\right) \implies \left(\frac{l_{s\tau}}{G_{s\tau}k_{s0}} > \frac{l'_{s\tau}}{G'_{s\tau}k'_{s0}}\right)$. Combined with $k_{s0} = \zeta_s k_s^u$, this entails that $(\mathcal{G}_{s\tau} < \mathcal{G}'_{s\tau}) \wedge \left(\frac{MC_{s\tau}}{MC_s^u} > \frac{MC'_{s\tau}}{MC_s^{u'}}\right)$ implies a contradiction. In point (ii), since $(m_s^u > m_s^{u'}) \implies \frac{\mu_s^u}{\mu_s^{u'}} > 1$, $\left(\frac{\mu_{s\tau}}{\mu_s^u} > \frac{\mu'_{s\tau}}{\mu_s^{u'}}\right)$ implies that $\mu_{s\tau} > \mu'_{s\tau}$, which is a contradiction with $(m_s^u > m_s^{u'}) \wedge (\mathcal{G}_{s\tau} < \mathcal{G}'_{s\tau})$.

The combination of point (i) and (ii) entails a contradiction with the supposition. Hence, its opposite must be true, which proves that $(m_s^u > m_s^{u'}) \wedge (G_{s\tau-1} > G'_{s\tau-1}) \implies (m_{s\tau} > m'_{s\tau})$.

Inductive Step B To examine capital growth, I update equation (32) to relative revenue at age τ:

$$\frac{\pi_{s\tau}}{a_{s-1}} = (\mu_{s\tau} - (1-\alpha))\frac{l_{s\tau}}{k_{s\tau}} G_{s\tau-1}\frac{w}{1-\alpha},$$

where $a_{s-1} = a_{s\tau-1}/G_{s\tau-1} = k_{s\tau}/G_{s\tau-1}$. I start by noting that $\mu_{s\tau} > \mu'_{s\tau}$, since $m_{s\tau} > m_{s\tau-1}$ and that $G_{s\tau-1} > G'_{s\tau}$ by assumption. Hence, the only component that remains to be analyzed is $\frac{l_{s\tau}}{k_{s\tau}}$, which I do in two cases.

- Suppose $\frac{l_{s\tau}}{k_{s\tau}} \geq \frac{l'_{s\tau}}{k'_{s\tau}}$. This case immediately implies that $(\frac{\pi_{s\tau}}{a_{s-1}} > \frac{\pi'_{s\tau}}{a'_{s-1}})$.
- Case (ii): suppose the firm, following the reaction function in Equation (19), chooses optimally to have $\frac{l_{s\tau}}{k_{s\tau}} < \frac{l'_{s\tau}}{k'_{s\tau}}$. Since the reaction function in Equation (19) maximizes the revenue

[27] The difference in demand elasticity between the firm in the two equilibria, implies that it is possible to have $\pi_{s0}\left(\frac{l_{s0}}{k_{s0}} = \frac{l'_{s0}}{k'_{s0}}\right) < \pi_{s0}\left(\frac{l_{s0}}{k_{s0}} < \frac{l'_{s0}}{k'_{s0}}\right)$, while $\pi'_{s0}\left(\frac{l'_{s0}}{k'_{s0}} = \frac{l_{s0}}{k_{s0}}\right) < \pi'_{s0}\left(\frac{l'_{s0}}{k'_{s0}} > \frac{l_{s0}}{k_{s0}}\right)$.

net of labor costs of the firm within period τ, and since $\frac{l_{s\tau}}{k_{s\tau}} \geq \frac{l'_{s\tau}}{k'_{s\tau}}$ is within the firm's choice set, setting $\frac{l_{s\tau}}{k_{s\tau}} < \frac{l'_{s\tau}}{k'_{s\tau}}$ implies that its revenue net of labor costs is weakly higher than under case (i), so $(\frac{\pi_{s0}}{a_{s-1}} > \frac{\pi'_{s0}}{a'_{s-1}})$ continues to hold. What happens in this case (ii), is that the firm optimally restricts output to maximize revenue, and this restricted output is optimal given that its demand is more inelastic.

This completes the demonstration of the inductive step. The combination of the first step and the inductive step together imply that Lemma 2 holds.

A.2 Proof on Markup Dispersion in Proposition 1

I examine how the markups of unconstrained firms $v = u$ or firms of age $v = \tau > 0$ behave relative to those of firms of age 0:

$$\frac{\partial \frac{\mu_{sv}}{\mu_{s0}}}{\partial M_s} = \frac{\partial \frac{\varepsilon(m_{sv})\varepsilon(m_{s0}) - \varepsilon(m_{sv})}{\varepsilon(m_{sv})\varepsilon(m_{s0}) - \varepsilon(m_{s0})}}{\partial M_s}$$

Working out the derivative and simplifying:

$$\frac{\partial \frac{\mu_{sv}}{\mu_{s0}}}{\partial M_s} = \frac{\varepsilon(m_{sv})(\varepsilon(m_{sv}) - 1)\frac{\partial \varepsilon(m_{s0})}{\partial m_{s0}}\frac{\partial m_{s0}}{\partial M_s} - \varepsilon(m_{s0})(\varepsilon(m_{s0}) - 1)\frac{\partial \varepsilon(m_{sv})}{\partial m_{sv}}\frac{\partial m_{sv}}{\partial M_s}}{(\varepsilon(m_{sv})\varepsilon(m_{s0}) - \varepsilon(m_{s0}))^2}$$

Plugging in the values for $\frac{\partial \varepsilon(m_{ist})}{\partial m_{ist}}$ for the Cournot demand elasticity:[28]

$$\frac{\partial \frac{\mu_{sv}}{\mu_{s0}}}{\partial M_s} = \frac{(1 - \frac{1}{\sigma})\frac{\varepsilon(m_{s0})(\varepsilon(m_{s0})-1)}{\varepsilon(m_{sv})^2}\frac{\partial m_{sv}}{\partial M_s} - (1 - \frac{1}{\sigma})\frac{\varepsilon(m_{sv})(\varepsilon(m_{sv})-1)}{\varepsilon(m_{s0})^2}\frac{\partial m_{s0}}{\partial M_s}}{(\varepsilon(m_{sv})\varepsilon(m_{s0}) - \varepsilon(m_{s0}))^2} < 0 \tag{33}$$

Here we have that $\frac{\varepsilon(m_{sv})(\varepsilon(m_{sv})-1)}{\varepsilon(m_{s0})^2} < \frac{\varepsilon(m_{s0})(\varepsilon(m_{s0})-1)}{\varepsilon(m_{sv})^2}$.

I first focus on constrained firms of age $\tau > 0$, where I consider two cases. First $\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} > 0$, and then its opposite.

- Case (i), suppose $\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} > 0$. Note that $\frac{y_{s\tau}}{y_{s0}} = \left(\frac{MC_{s0}}{MC_{s\tau}}\frac{\mu_{s0}}{\mu_{s\tau}}\right)^\sigma$ and $\frac{\partial G_{s\tau}}{\partial M_s} < 0$. There are then two subcases:
  - $\frac{\partial l_{s\tau}/k_{s0}}{\partial M_s} < 0$, in which case $\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} < 0$ since $\frac{\partial G_{s\tau}}{\partial M_s} < 0$, which would entail a contradiction with the supposition, so it cannot hold.
  - $\frac{\partial l_{s\tau}/k_{s0}}{\partial M_s} \geq 0$, which entails $\frac{\partial MC_{s\tau}}{\partial M_s} < 0$. In that case, having $\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} > 0$ requires that

$$\frac{\partial \frac{\mu_{s\tau}}{\mu_{s0}}}{\partial M_s} < 0.$$

- In case (ii), suppose $\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} \leq 0$. It is then immediate that $\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} \leq 0 \iff \frac{\partial \frac{m_{s\tau}}{m_{s0}}}{\partial M_s} \leq 0$, where $\frac{\partial \frac{m_{s\tau}}{m_{s0}}}{\partial M_s} \leq 0 \implies \left(\frac{\partial m_{s\tau}}{\partial M_s}m_{s0} \leq \frac{\partial m_{s0}}{\partial M_s}m_{s\tau}\right)$. The implied component can be rewritten as $\left(\frac{\partial m_{s\tau}}{\partial M_s} \leq \frac{\partial m_{s0}}{\partial M_s}\frac{m_{s\tau}}{m_{s0}}\right)$. Since $m_{s0} < m_{s\tau}$ and $\frac{\partial m_{s0}}{\partial M_s}, \frac{\partial m_{s\tau}}{\partial M_s} < 0$ from Lemma 2, in this second case it holds that:

$$\frac{\partial \frac{y_{s\tau}}{y_{s0}}}{\partial M_s} \implies \frac{\partial m_{s\tau}}{\partial M_s} < \frac{\partial m_{s0}}{\partial M_s} < 0.$$

[28] It is straightforward to verify that the following result also holds for the Bertrand demand elasticity.

Given Equation (33), this has the same implication as in the first case, namely that:

$$\frac{\partial \frac{\mu_{s\tau}}{\mu_{s0}}}{\partial M_s} < 0.$$

The remaining question is how the relative markups of unconstrained firms behave. Here, note that $m^u_s = \mathcal{G}_{s0} m_{s0}$, such that $\frac{\partial m^u_s}{\partial M_s} = \mathcal{G}_{s0} \frac{\partial m_{s0}}{\partial M_s} + m_{s0} \frac{\partial \mathcal{G}_{s0}}{\partial M_s}$.

Now, suppose $\frac{\partial \mu^u_s / \mu_{s0}}{\partial M_s} \geq 0$. Given that $\mathcal{G}_{s0} = \left(\frac{MC_{s0} \mu_{s0}}{MC^u_s \mu^u_s} \right)^{\sigma - 1}$, this supposition implies $\frac{\partial \mathcal{G}_{s0}}{\partial M_s} \leq 0$, because an increase in MC_{s0}/MC^u_s - i.e. the only way $\mathcal{G}_{s0}$ could increase - would require l_{s0}/k_{s0} to increase which implies $\mathcal{G}_{s0}$ falls.

Then, since $\frac{\partial \mathcal{G}_{s0}}{\partial M_s} \leq 0$, $\frac{\partial m^u_s}{\partial M_s} < \frac{\partial m_{s0}}{\partial M_s}$, which from equation (33) implies that $\frac{\partial \frac{\mu_{sH}}{\mu_{s0}}}{\partial M_s} < 0$. Hence, the supposition that $\frac{\partial \frac{\mu_{sH}}{\mu_{s0}}}{\partial M_s} \geq 0$ entails a contradiction, and therefore its opposite must be true:

$$\frac{\partial \frac{\mu^u_s}{\mu_{s0}}}{\partial M_s} < 0.$$

Appendix B Data description

B.1 Annual Survey of Industries

In the empirical analysis, I employ plant-level data from the Annual Survey of Industries (ASI), obtained from India's Ministry of Statistics and Programme Implementation (MOSPI). I use data from each survey year for the accounting years 1990-91 till 2011-2012. As explained in Section 3.1, the ASI sampling design has been updated from time to time. Most importantly, for most years the census scheme includes all plants with 100 or more employees. The exceptions are the years 1997-98 till 1999-00, when this threshold was 200 workers instead. Further details on the exact sampling design for each year can be found on the MOSPI website.[29]

For the cleaning of the data, I follow a procedure that is highly similar to the data cleaning in Allcott et al. (2016). First, I correct observations in the 1993-94 to 1997-98 survey years, whose values have been provided in "pre-multiplied" form by MOSPI.[30] Second, I drop duplicate observations – a small but non-trivial issue in the early years of the sample – and observations missing state identifiers. Third, I restrict the sample to all plants (i) in the manufacturing sector, (ii) who are listed as open, and (iii) who have non-missing and positive values for three critical variables, namely the logarithms of revenue, capital and labor cost. Fourth, I drop observations whose labor cost or material cost exceeds 250 percent of their revenue.

Two of the main plant-level variables of interest in the analysis are defined as ratios: the labor cost share, and the ratio of revenue to capital. Since these are ratios, price inflation is an irrelevant concern in their computation. When computing these ratios, I therefore measure both the numerator and denominator in nominal terms. However, a third variable of interest - the capital growth rate - is sensitive to inflation in its measurement and hence I calculate it after deflating the book value of capital with the capital deflator from the Indian Handbook of Industrial Statistics, setting 2004 as the base year. To mitigate the influence of outliers, I winsorize all three variables - the labor cost share, the ratio of revenue to capital, and the capital growth rate - at the 1st and 99th percentile, within each sector-year cell, with sectors defined at the 2-digit level.

In the analysis, I define sectors at the 3-digit level, using India's 1987 National Industrial Classification (NIC). There are 191 separate 3-digit sectors in my data. I also ensure that plants are assigned to a consistent geographic region (state or union territory) over all years. To make the definitions of regions consistent over time, I employ the concordance provided by the Indian Statistical Office. In addition, I take into account the creation of Jharkhand, Chhattisgarh and Uttaranchal in 2001 from Bihar, Madhya Pradesh and Uttar Pradesh, respectively. I assign these newly created states the state code from the state they separated from. There are 29 separate geographic regions in my data.

B.2 Data on dereservation

As described in Section 3.1, I employ the concordance between SSI product codes, used to determine dereservation status, and ASICC product codes, used in the ASI data, available from Martin

[29]The sampling designs for all years prior to 1998-99 is described on http://mospi.nic.in/salient-features-sampling-designs-asi-1973-74-asi-1998-99, and the designs from 1998-99 onward on http://www.csoisw.gov.in/cms/cms/files/554.pdf. Both links were retrieved on February 22, 2019.

[30]For further detail on this pre-multiplication, see for instance the documentation on MOSPI's microdata catalogue http://mail.mospi.gov.in/index.php/catalog/75#page=sampling&tab=study-desc (retrieved on February 22, 2019.)

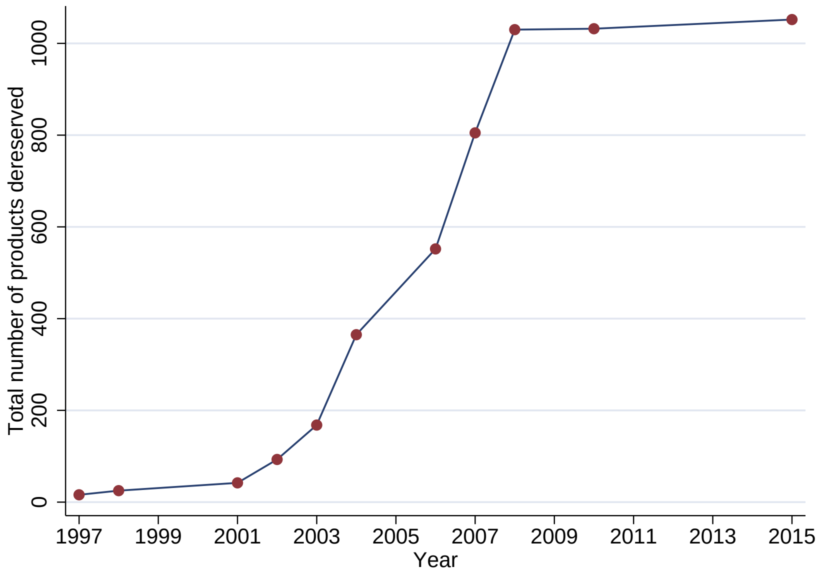
et al. (2017). A substantial subset of the ASICC codes is relatively broad. For this reason, Martin et al. (2017) construct a first set of matches between the SSI product codes and pairs of an ASICC code and a 5-digit NIC code. Remaining product codes are simple matches between the SSI and ASICC code.

I cross-checked the Martin et al. (2017) dereservation list with the list from Tewari and Wilde (2017), which Ishani Tewari generously made available to me. The two lists correspond very closely, except for seven products that are included in Tewari and Wilde (2017), but not in Martin et al. (2017). These seven products are also included in the dereservation notifications on the website of the Indian Ministry for Micro, Small and Medium Enterprises (referenced in footnote 8). I therefore add these omitted products to the product-level concordance.

Starting 2010, the ASI changes its product classification from ASICC (A Standard Industrial Commodity Classification) to NPCMS (National Product Classification for Manufacturing Sector). Since the concordance between the ASICC and NPCMS classifications is not one-to-one, I only consider products that were dereserved prior to 2010. This covers 97.9% of reserved products. An additional complication is that the ASI data has incomplete product coverage for the years 1998 and 1999. A potential concern is that the covered products in those years are a non-randomly selected sample, which could introduce bias into the estimation results. To address this issue, I implement robustness checks where I restrict the sample to incumbent plants with dereservation years after 1999 and before 2010 (see Appendix Figure C.6 and Table C.2).

Appendix C Supplementary tables and figures

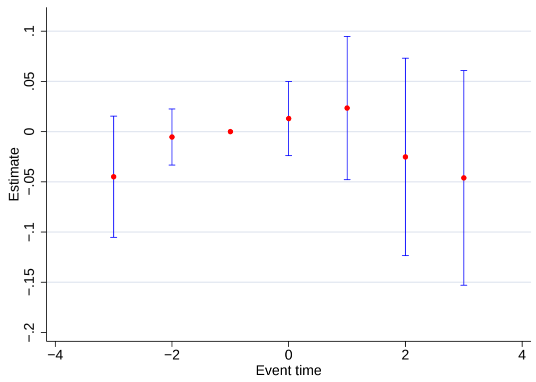
Figure C.1: Timing of dereservation



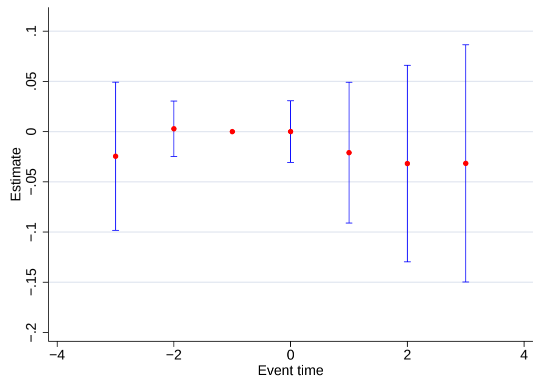
The figure displays the cumulative number of products dereserved in a given year, from the start of the dereservation reform in 1997 till the completion of dereservation in 2015. Each dot indicates a year when products were dereserved.

Figure C.2: Absence of pre-trends for dereserved plants

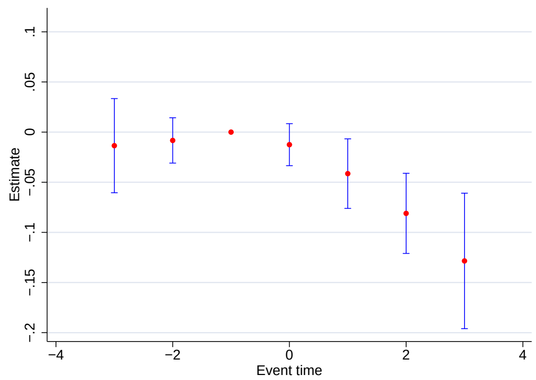
(a) Revenue



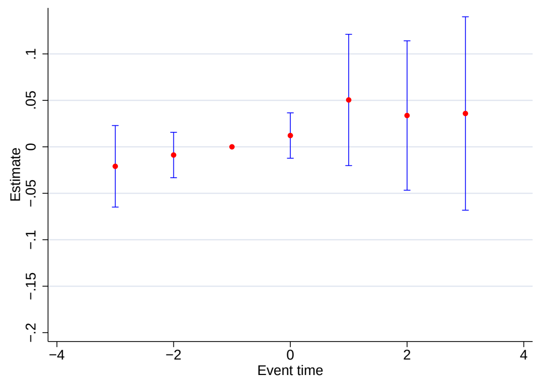
(b) Capital



(c) Number of employees

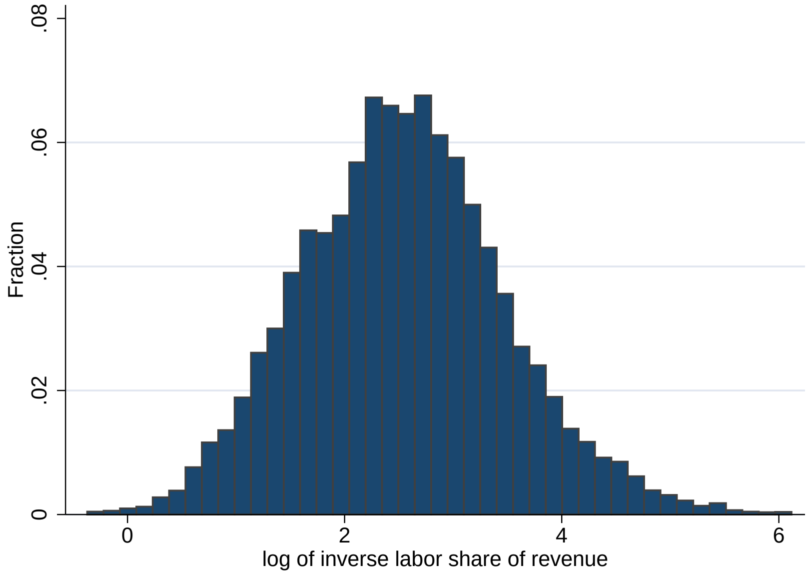


(d) Labor cost



The figure displays the coefficients and 95% confidence intervals for the β_τ coefficients from the following event-study regressions: $\ln y_{it} = \gamma_i + \nu_t + \sum_{\tau=-3}^{3} \beta_\tau 1[t = e_{it} + \tau] + \varepsilon_{it}$, where γ_i is a plant fixed-effect and ν_t is a year fixed effect. The dependent variable y_{it} is revenue, capital, number of employees, or total labor cost, for panels a,b,c and d respectively. The regression is estimated on a balanced sample of between 20913 to 20937 incumbent plants, depending on the outcome variable. I define the time at which the first product of plant i is dereserved as e_{it}, and impose the normalization that $\beta_{-1} = 0$. Since dereservation status is defined at the product level, I also cluster standard errors at that level. Revenue, capital and labor cost are deflated using the deflators of the Indian Handbook of Industrial Statistics, with 2004 as the base year.

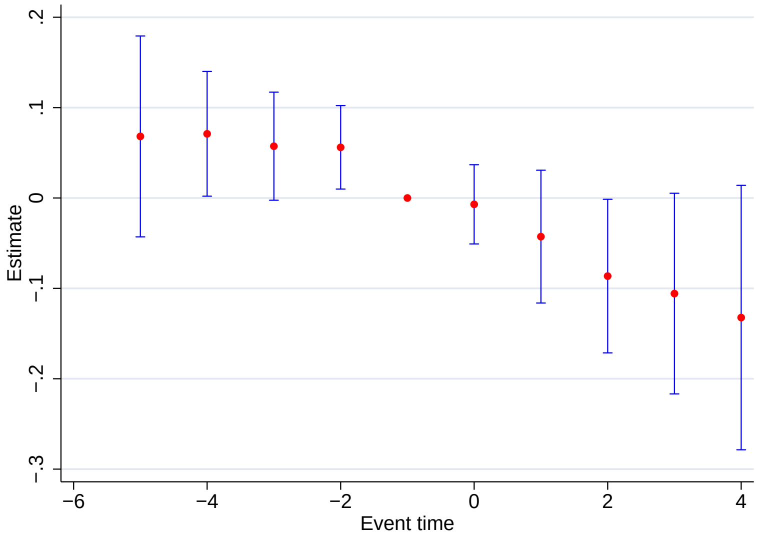
Figure C.3: Inverse labor share among plants in the dereservation event study



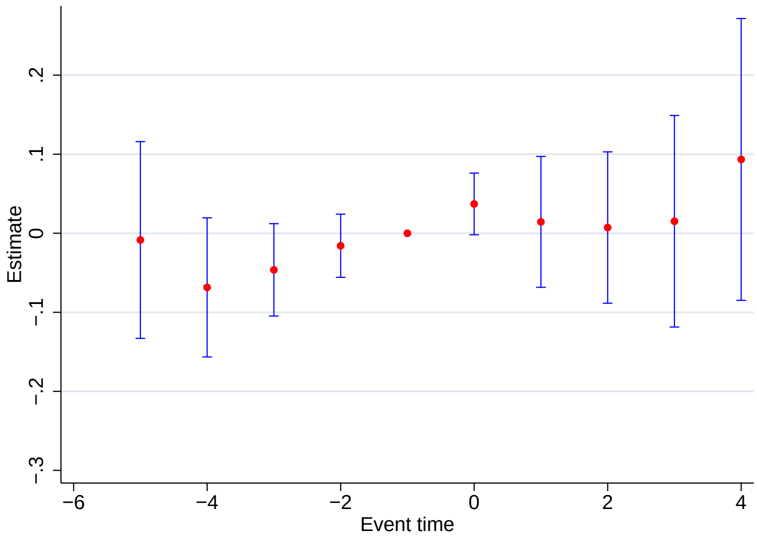
The figure displays the histogram for $\ln(S_{it}/w_t L_{it})$ among plants in the dereservation event study. This includes all plants who were incumbents producing reserved products, and who were observed each of the three periods before and three periods after their first product was dereserved. The variable $\ln(S_{it}/w_t L_{it})$ is driving the within-plant variation in the markups in the event study in Figure 1, since the output elasticity is absorbed in a plant fixed-effect.

Figure C.4: Event study of dereservation over a longer time horizon

(a) Incumbents with above-median initial markup



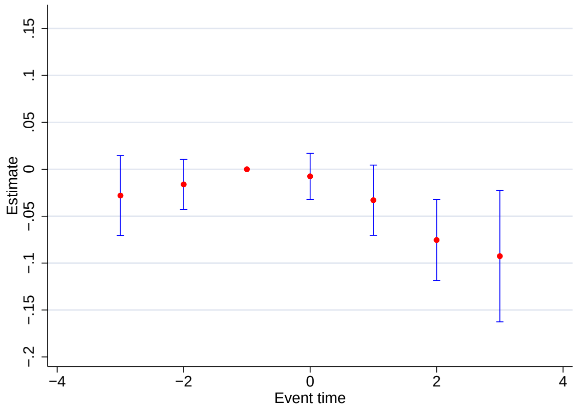
(b) Incumbents with below-median initial markup



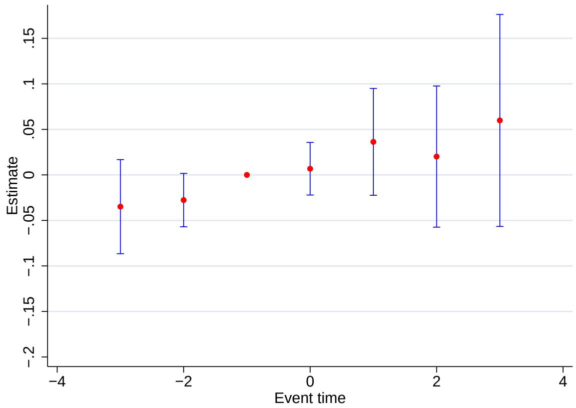
The figure displays the coefficients and 95% confidence intervals for the β_τ coefficients from the following event-study regression: $\ln \mu_{it} = \gamma_i + \nu_t + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{it} + \tau] + \varepsilon_{it}$, where γ_i is a plant fixed-effect and ν_t is a year fixed effect. I define the time at which a plant's first product of plant i is dereserved as e_{it}, and impose the normalization that $\beta_{-1} = 0$. Since dereservation status is defined at the product level, I also cluster standard errors at that level. I restrict the sample to a balanced sample of incumbent plants that are observed at least five years before and four years after they are dereserved. Panel (a) shows results for plants with initial markups weakly above the median initial markup, and Panel (b) for the other plants. The initial markup is an average over event times $\tau = -4$ till $\tau = -2$, and the median initial markup is then set after taking out sector and year fixed effects.

Figure C.5: Event study of dereservation for incumbents with high initial markups

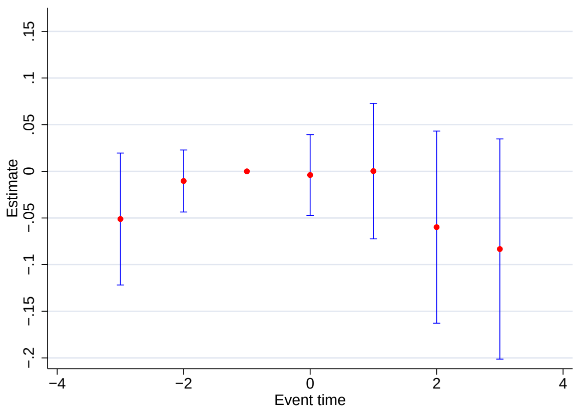
(a) Number of Employees



(b) Labor Cost



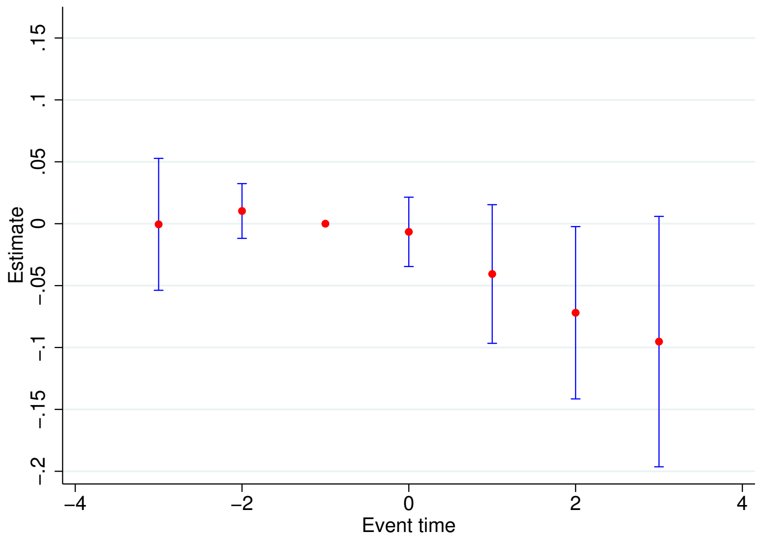
(c) Revenue



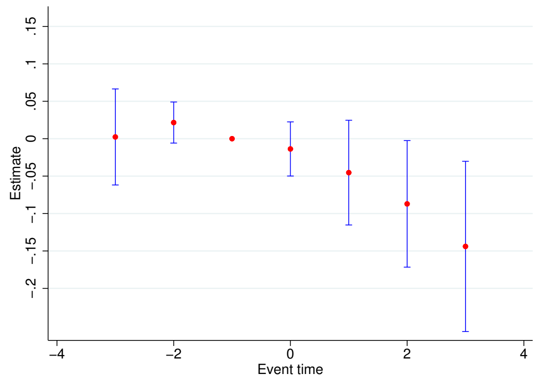
The figure displays the coefficients and 95% confidence intervals for the β_τ coefficients from the following event-study regressions: $\ln y_{it} = \gamma_i + \nu_t + \sum_{\tau=-3}^{3} \beta_\tau 1[t = e_{it} + \tau] + \varepsilon_{it}$, where γ_i is a plant fixed-effect and ν_t is a year fixed effect. The dependent variable y_{it} is number of employees, total labor cost, or revenue for panels a, b and c respectively. The regression is estimated on a balanced sample of plants above the median initial markup, as defined in Figure 1. I define the time at which the first product of plant i is dereserved as e_{it}, and impose the normalization that $\beta_{-1} = 0$. Since dereservation status is defined at the product level, I also cluster standard errors at that level. Revenue and labor cost are deflated using the deflators of the Indian Handbook of Industrial Statistics, with 2004 as the base year.

Figure C.6: Robustness test for the event study for the impact of dereservation on markups

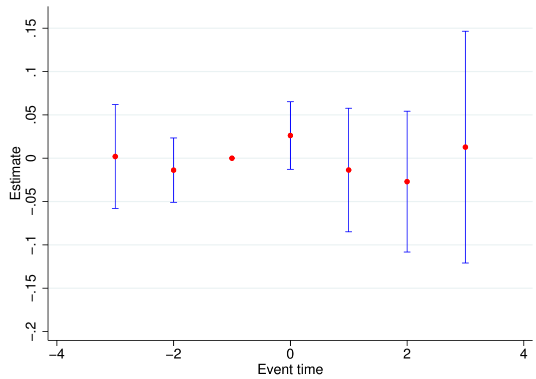
(a) Full sample of incumbents



(b) Incumbents with above-median initial markup

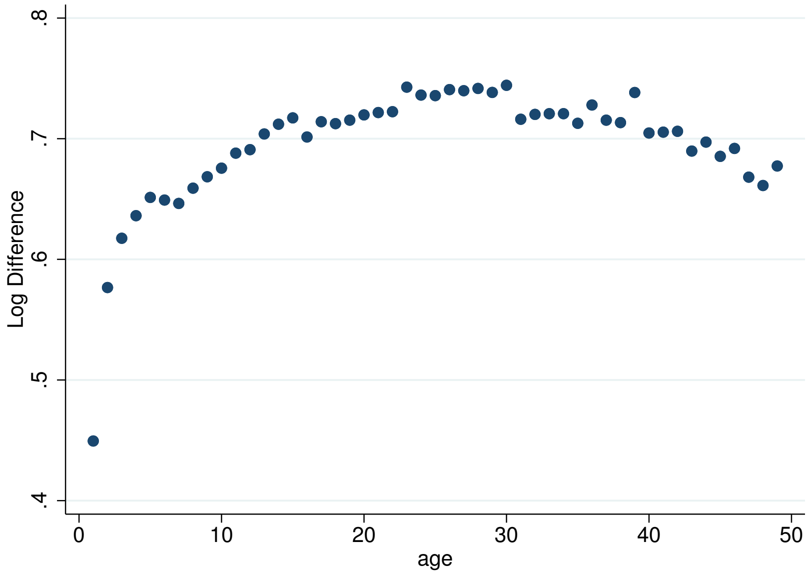


(c) Incumbents with below-median initial markup



This figure implements a robustness test for the results in Figure 1, by restricting the sample to incumbent plants dereserved after 1999, which is motivated by incomplete product coverage in the data before 1999. The figure displays the coefficients and 95% confidence intervals for the β_τ coefficients from the following event-study regression: $\ln \mu_{it} = \gamma_i + \nu_t + \sum_{\tau=-3}^{3} \beta_\tau 1[t = e_{it} + \tau] + \varepsilon_{it}$, where γ_i is a plant fixed effect and ν_t is a year fixed effect. I define the time at which a plant's first product of plant i is dereserved as e_{it}, and I impose the normalization that $\beta_{-1} = 0$. Since dereservation status is defined at the product-level, I also cluster standard errors at that level. I restrict the sample to a balanced sample of incumbent plants that are observed at least three years before and after they are dereserved. Panel (a) shows results for the full sample of incumbent plants. Panel (b) displays results for plants with initial markups weakly above the median initial markup, and Panel (c) for the other plants. The initial markup is an average over event times $\tau = -3$ and $\tau = -2$, and the median initial markup is then set after taking out sector and year fixed effects. Since the ASI's product classification changes after 2010, this analysis is performed on plants who are dereserved after 1999 and before 2010.

Figure C.7: Log size difference between plants of different age



The figure displays the estimated β_τ coefficients from the regression $\ln v_{it} = \gamma_t + \sum_\tau \beta_\tau age_{i\tau} + \varepsilon_{it}$, where v_{it} is revenue, γ_t are year fixed effects and $age_{i\tau}$ are indicator variables for whether plant i is of age τ. The indicator for $\tau = 0$ is the left out category, and for convenience, only coefficients up until age 50 are displayed.

Table C.1: Summary statistics

	Full Panel	Dereservation Incumbents
Age of Plant	17	18
	(17)	(15)
Number of workers	174	190
	(801)	(834)
Capital Growth	.058	.054
	(.38)	(.37)
Revenue to capital ratio	712	369
	(39540)	(29898)
Labor cost share of gross revenue	.13	.13
	(.17)	(.16)
Observations	697183	163519

The table shows average values, with standard deviations in parentheses, for the variables in each column. The full panel covers all plants with non-missing capital growth, included in the estimation of column 1 of Table 2. The sample of dereservation incumbents covers all plants that are incumbents in the dereservation analysis with non-missing capital growth, included in the estimation of column 1 of Table 1. I include the revenue to capital ratio and the labor cost share in this table because these two variables are central to the measurement of MRPK, central to the analysis in Appendix D, and markups, respectively.

Table C.2: Young plants' capital growth for dereservation years 2000-2009

	Capital growth $g(k)_{it}$		
	(1)	(2)	(3)
$Deres_{it-1} * 1(age_{it} \leq 5)$	-0.029***		
	(0.009)		
$Deres_{it-1} * 1(age_{it} < 10)$		-0.018***	
		(0.005)	
$Deres_{it} * [-\ln age_{it}]$			-0.009***
			(0.003)
$1(age_{it} \leq 5)$	0.025***		
	(0.010)		
$1(age_{it} < 10)$		0.010	
		(0.006)	
$-\ln age_{it}$			0.002
			(0.004)
$Deres_{it-1}$	-0.001	-0.002	-0.023**
	(0.006)	(0.006)	(0.010)
State FE, Sector FE, Year FE	Yes	Yes	Yes
Observations	124832	124832	122989

Robustness analysis for Table 1. Since product coverage in the ASI is incomplete for the years 1998 and 1999, and since product classification changes after 2009, the estimation in this table is restricted to the incumbent plants that are dereserved during the years 2000-2009. Since dereservation status is defined at the product level, I also cluster standard errors at that level. Since I restrict the analysis to a subsample of dereservation years, I do not estimate the full sample analysis (equation (27)) in this table. Standard errors in parentheses. $^{*}p < 0.1; ^{**}p < 0.05; ^{***}p < 0.01$.

Appendix D Adapting the framework to mature firms

The main theoretical and empirical analysis in the paper focuses on young firms. Yet, the central insights from the model can also be derived in a setting with "full-grown" firms, namely by introducing volatility in individual plant's productivity. This is done in this appendix section, which then proceeds by testing the prediction of this model. The results further corroborate that competition slows down capital convergence, also for mature firms.

D.1 A model with productivity volatility

Here I present a shortened version of the model with productivity volatility, the full version of which is available in the 2019 version of this paper. The theory is identical to the model in Section 2, except on the following dimensions. First, instead of having birth and death of firms in the model, all firms are equally old and infinitely lived. Second, firm's individual productivity follows a stochastic process over two values: z_{sL} or z_{sH}, with $z_{sL} < z_{sH}$. Under the strong assumptions that firms are unable to save and have no access to external finance, the distribution of capital takes on the following form in steady state:

Lemma D.1. *In steady state, the joint distribution of capital and productivity within a sector is as follows:*

- *all low-productivity firms are unconstrained: if $z_{ist} = z_{sL}$, then $k_{ist} = k_{sL}$*
- *high-productivity firms can be constrained or unconstrained, depending on the number of periods τ since their most recent productivity shock, and constrained firms invest all their wealth into capital growth: if $z_{ist} = z_{sH}$, then $\forall i$ with $\tau = t - v$, where $v \equiv \max r$ s.t. $z_{isr+1} = z_{sH} \& z_{isr} = z_{sL}$:*
  - *if constrained, then $k_{ist} = G_{s\tau} k_{sL}$, with $G_{s\tau} \equiv \Pi_{r=s}^{s+\tau} \left(\frac{\pi_{sv}}{P^F k_{sv}} + 1 - \delta \right)$*
  - *if unconstrained, then $k_{ist} = k_{sH}$*

Given this capital distribution, and with τ defined as in the above lemma, the following proposition holds.

Proposition D.1. *For any $M_s' > M_s$, and for unconstrained firm-types L, H, and for constrained firms in bin $\tau > 0$:*

- *Markup levels fall with M_s:*

$$\mu'_{sL} < \mu_{sL} \, ; \, \mu'_{sH} < \mu_{sH} \, ; \, \mu'_{s\tau} < \mu_{s\tau}$$

- *Markup dispersion falls with M_s:*

$$\frac{\mu'_{sH}}{\mu'_{sL}} < \frac{\mu_{sH}}{\mu_{sL}} \, ; \, \frac{\mu'_{s\tau}}{\mu'_{sL}} \leq \frac{\mu_{s\tau}}{\mu_{sL}}$$

- *Capital growth rates for all financially constrained firms fall with M_s:*

$$G'_{s\tau} < G_{s\tau}.$$

D.2 Dereservation and MRPK convergnece

This result implies that all plants, regardless of age, can exhibit capital convergence, and that competition slows down this capital convergence. To test this, first I need a measure of capital convergence. Recall that in the model, firms optimally choose to grow their capital stock in response to positive productivity shocks until they reach their optimal, unconstrained level of capital. The empirical challenge is that the optimal level of capital is unobserved. Interestingly, while a positive productivity shock leads to a first-order increase in the optimal level of capital, the change in the optimal marginal revenue product of capital (MRPK) is second order. As a result, it is feasible to find valid proxies for optimal MRPK. For this reason, and inspired by Asker et al. (2014), I focus on convergence in MRPK in my empirical analysis of capital convergence. When a firm is financially constrained, its actual $MRPK_{it}$ will be above its optimal, unconstrained MRPK, denoted by $MRPK^*_{it}$. Since $MRPK_{it}$ is a strictly monotonic function of a firm's capital level, when capital convergence in the model slows down, convergence in terms of MRPK also slows down.

As in Asker et al. (2014), I measure MRPK in logs after assuming Cobb Douglas production functions:

$$MRPK_{it} = \ln \alpha_i^K + s_{it} - k_{it},$$

where α_i^K is the output elasticity of capital, s_{it} is log revenue, and k_{it} is log capital. To allow for maximal cross-plant heterogeneity in α_i^K, I absorb this output elasticity in a plant fixed effect in the regression analysis. Hence, within-plant variation in MRPK will be driven by the log ratio of revenue to capital. To examine the sensitivity of the results to the precise measurement of MRPK, I also measure it using log value-added instead of log revenue. Since the measurement in logarithms requires value added to be positive, this leading robustness check is also performed on a more restrictive selection of "well-performing" plants.

As in the main analysis in this paper, I then employ two variables to measure changes in competition: the dereservation reform and the median markup at the state-sector level. First, I focus on the dereservation reform, where I apply the following autoregression framework:

$$\begin{aligned} MRPK_{it} = \gamma_i + \nu_t + \beta_1 Deres_{it-1} + \rho_0 MRPK_{it-1} \\ + \rho_1 MRPK_{it-1} * Deres_{it-1} + \beta_2 \ln age_{it} + \varepsilon_{irt} \end{aligned} \tag{34}$$

where γ_i and ν_t are plant and year fixed effects respectively, and $Deres_{it-1}$ indicates if a first product of plant i has been dereserved in period $t-1$ or earlier. I estimate equation (34) both on a sample with only incumbents, as well as on the full sample of plants. In the latter setup, I can control for economic shocks at the region-sector-year level, which is impossible when restricting the sample to incumbents, due to collinearity issues.[31]

The main coefficient of interest is ρ_1, which estimates how the speed of convergence to $MRPK_{it}^*$ changes after dereservation. To understand the estimation strategy, consider the case when $\rho_0 = \rho_1 = 0$. In that case, plants exhibit immediate convergence to $MRPK_{it}^* \equiv E[MRPK_{it}|(\rho_0 = \rho_1 = 0)]$, regardless of $MRPK_{it-1}$. In practice however, the average plant experiences a delayed adjustment to $MRPK_{it}^*$, with $\rho_0 > 0$. The closer ρ_0 is to unity, the slower convergence to $MRPK_{it}^*$. Crucially, when $\rho_0 > 0$, then $\rho_1 > 0$ indicates that the speed of MRPK convergence slows down after dereservation. In equation (34), I proxy for $MRPK_{it}^*$ with $\gamma_i + \nu_t + \beta_1 Deres_{it-1} + \beta_2 \ln age_{it}$, though the findings on ρ_1 are robust to the exact choice of proxy. Still, it is important to include the indicator variable for dereservation status. After all, dereservation implies the removal of the size restriction on capital, which may have implications for the optimal capital share. Including the $Deres_{it-1}$ indicator, entails that $MRPK_{it}^*$ can adjust accordingly after the dereservation reform.

[31] Collinearity issues arise from small numbers of incumbent plants in many region-sector-year observations. For the estimation on the full sample, I distinguish between three types of plants. A first type is the incumbent plant, defined above. A second type is the "entrant" plant, which after dereservation starts producing a previously reserved product. The third type of plant - labeled as "outsider" - includes all remaining plants. For this full sample of plants, I employ the following estimation specification:

$$\begin{aligned} MRPK_{irst} = \gamma_{irs} + \nu_{rst} + \beta_1 Deres_{irst-1} + \beta_2 Deres_{irst-1} * entrant_{irs} \\ + \rho_0 MRPK_{irst-1} + \rho_2 MRPK_{irst-1} * entrant_{irs} + \rho_1 MRPK_{irst-1} * outsider_{irs} \\ + \rho_3 MRPK_{irst-1} * Deres_{irst-1} + \rho_4 MRPK_{irst-1} * Deres_{irst-1} * entrant_{irs} \\ + \beta_3 X_{irst} + \varepsilon_{irst} \end{aligned} \tag{35}$$

Here, γ_{irs} is a plant fixed-effect, $entrant_{irs}$ and $outsider_{irs}$ are indicators for plant i being entrants or outsiders. Next, ν_{rst} is a region-sector-year fixed effect that absorbs local economic shocks. While lengthy, the above specification is still intuitive. The top row is a standard difference-in-difference framework, where I allow for different $MRPK_{it}^*$ levels post dereservation for incumbents and entrants. The middle row estimates convergence speeds prior to dereservation, allowing for different speeds of convergence for incumbents, entrants and outsiders. The third row then estimates how speeds of convergence change after dereservation, where ρ_3 - the coefficient of interest - estimates how speed of converges changes for incumbent firms.

Table D.1: Dereservation and MRPK convergence

	$MRPK_{it}$ - Gross Revenue (GR)		$MRPK_{it}$ - Value Added (VA)	
	(1)	(2)	(3)	(4)
$Deres_{it-1}$	-0.118**	-0.061*	-0.165**	-0.029
	(0.023)	(0.027)	(0.018)	(0.021)
$MRPK_{it-1}(GR)$	0.439**	0.369**		
	(0.016)	(0.020)		
$MRPK_{it-1}(GR) * Deres_{it-1}$	0.031**	0.034*		
	(0.011)	(0.014)		
$MRPK_{it-1}(VA)$			0.306**	0.244**
			(0.012)	(0.020)
$MRPK_{it-1}(VA) * Deres_{it-1}$			0.032**	0.032+
			(0.011)	(0.017)
Plant Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	–	Yes	–
State-sector-year Fixed Effects	No	Yes	No	Yes
Observations	62482	186688	56482	168314

In specifications 1 and 2, MRPK is measured based on gross revenue, and based on value added in specifications 3 and 4. Since there is a change in product codes after 2010, specifications 1 and 3 estimate equation (34) on a sample restricted to a sample of incumbent plants dereserved before 2010, while specifications 2 and 4 estimate equation (35) on the full sample for all sample years before 2010. All specifications control for the logarithm of a plant's age. Standard errors, in parentheses, are clustered at the product level, which is the level at which dereservation status is defined. $^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$.

The empirical results are in line with the theoretical predictions of the model (see Table D.1). First, I find that there is indeed convergence to $MRPK^*_{it}$, since ρ_0 is significantly below 1, but this convergence is not immediate as ρ_0 is also significantly above 0. This is consistent with the speed of convergence being limited by the presence of financial constraints. The point estimates for ρ_0 are below 0.5, which implies that convergence to $MRPK^*_{it}$ is relatively fast. Hence, the proxy for $MRPK^*_{it}$ appears to be empirically valid. Most importantly, all coefficients on the interaction of dereservation with $MRPK_{it-1}$ are positive - as predicted by the theory - and statistically significant. The estimated magnitude of the effect of dereservation is modest but economically meaningful. For specifications 1 and 2 specifically, dereservation increases the half-life of the autoregressive process by 9% and 9.7% respectively.[32]

One concern with this estimation procedure arises from the downward bias on autoregression coefficients described by Nickell (1981). While in principle it may be possible to obtain consistent estimates using GMM methods, these econometric approaches come with their own pitfalls (Roodman, 2009). Note however that my primary objective here is to provide qualitative support for the model's predictions. First then, observe that my panel spans the period 1990-2011, so the downward bias, which is of order $1/T$, becomes small. Second and most importantly, the downward bias on ρ_1 works against finding evidence for dereservation slowing down MRPK convergence.

As mentioned, my autoregression framework is inspired by the analysis on capital convergence in Asker et al. (2014). They employ this specification to show that heterogeneity in MRPK

[32] The following formula, which is derived from the AR(1) convergence process, computes the percentage increase in the half-life: $\frac{\log(0.5)/\log(\rho_0+\rho_1)}{\log(0.5)/\log(\rho_0)}$.

can be driven by delayed adjustment to productivity shocks. Hence, the purpose of their empirical analysis is very closely related to the one in this paper. The main difference between their setup and mine, is that they assume that the delayed adjustment to productivity shocks is driven by capital adjustment costs. However, as implied by my model, such delayed adjustment can also be caused by financial constraints. In my empirical setting, financial constraints are a much more likely driver of the results than adjustment costs. After all, I have provided evidence that plants' markups indeed fall after dereservation, which directly affects their retained earnings. Moreover, MRPK convergence slows down after dereservation. From the perspective of adjustment costs, this would require increases in adjustment costs to coincide with the timing of dereservation. Finally, in the next section I extend the analysis of MRPK convergence to the full panel of plants employing a different measure for competition. It is again unlikely that this measure for competition is correlated with the severity of adjustment costs.

D.3 Competition and MRPK convergence in the full panel

I again strengthen the external validity of the MRPK convergence results by extending the analysis to the full panel of plants, instead of focusing only on incumbent plants whose products become dereserved. To this end, I return to employing the median markup, with the markup measured as in equation (28), as my measure of competition. Specifically, I update the autoregression framework from specification (34) in the following way:

$$\begin{aligned} MRPK_{irst} = \alpha_{irs} + \gamma_{rst} + \beta \ln age_{irst} + \rho_0 MRPK_{irst-1} \\ + \rho_1 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] + \varepsilon_{irst} \end{aligned} \tag{36}$$

As before, the main coefficient of interest is ρ_1. This coefficient estimates how the speed of convergence changes as a function of $Median_{rst-1}[\ln \mu_{irst-1}]$. Recall that when $\rho_0 = \rho_1 = 0$, plants exhibit immediate convergence to the empirical proxy for $MRPK^*_{irst}$, regardless of $MRPK_{irst-1}$. In practice, plants experience delayed adjustment to $MRPK^*_{irst}$. The theoretical prediction is then that $\rho_1 < 0$, as this implies that the speed of MRPK convergence increases with $Median_{rst-1}[\ln \mu_{irst-1}]$. I cluster standard errors at the sector level in the estimation.

The estimation results (see Table D.2) are in line with the results from the analysis of dereservation. First, across all specifications, MRPK converges strongly to the empirical proxy for $MRPK^*_{irst}$ ($\rho < 1$), but this convergence is not immediate ($\rho > 0$). Second, for baseline specification (36), the speed of convergence always increases with the median markup. Specifically, the coefficient on ρ_1 is always negative and strongly statistically significant in three of the four specifications (see columns 2, 5, and 6). This confirms the qualitative prediction of the model that the speed of convergence slows down with competition. The magnitude of this effect is modest but economically meaningful, as in the case of dereservation. As an example, in specifications 2 and 6, an increase in the median markup by two standard deviations, decreases the half-life of MRPK convergence by 4.8% and 13.4% respectively.[33]

In a final estimation, I examine the role of financial dependence in the setting of MRPK convergence, by augmenting the earlier specification to allow for heterogeneous effects along financial

[33] Given the presence of the Nickell bias in this autocorrelation framework, it is also noteworthy that the predicted sign for ρ_1 switches from the dereservation setting to the full panel. Despite the potential downward bias on this coefficient, the estimation results pick up this sign switch.

dependence:

$$\begin{aligned} MRPK_{irst} &= \alpha_{irs} + \gamma_{rst} + \beta \ln age_{irst} + \rho_0 MRPK_{irst-1} \\ &\quad + \rho_1 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] + \rho_2 MRPK_{irst-1} * Fin\ Dep_s \\ &\quad + \rho_3 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] * Fin\ Dep_s + \varepsilon_{irst} \end{aligned} \tag{37}$$

For this specification, the expectation is that $\rho_3 < 0$, as a decrease in competition would speed up convergence more for plants in sectors with higher levels of financial dependence.

Here, the results are also in line with expectations (see columns 3,4,7,8). First note that the coefficient on $MRPK_{irst-1} * Fin\ Dep_s$ is always positive, which is consistent with MRPK convergence being slower in more financially dependent sectors. More importantly, the coefficient ρ_3, estimated on the triple interaction term, is always significantly negative. These results imply that the median markup speeds up MRPK convergence more in sectors with higher financial dependence. Consider for instance the industry producing electric machinery, which has a relatively high level of financial dependence at $Fin\ Dep_s = 0.77$. For this sector, an increase in the median markup by two standard deviations, decreases the half-life of MRPK convergence by 12.5% or 21.8%, according to specifications 4 and 8 respectively.

Table D.2: Competition and Speed of MRPK Convergence

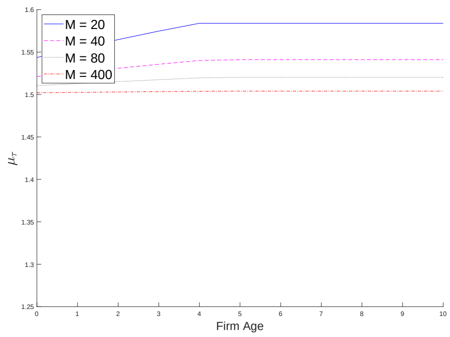
	$MRPK_{irst}$ (Gross Revenue (GR))				$MRPK_{irst}$ (Value added (VA))			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$MRPK_{irst-1}(GR)$	0.392^{**}	0.372^{**}	0.394^{**}	0.367^{**}				
	(0.013)	(0.014)	(0.015)	(0.019)				
$MRPK_{irst-1}$ (GR) * $Median_{rst-1}[\ln \mu_{irst-1}]$	-0.000	-0.009^{*}	0.014^{**}	-0.000				
	(0.003)	(0.004)	(0.005)	(0.005)				
$MRPK_{irst-1}$ (GR) * $Fin\ Dep_s$			0.005	0.012				
			(0.013)	(0.017)				
$MRPK_{irst-1}$ (GR)* $Median_{rst}[\ln(\mu_{irst}] * Fin\ Dep_s$			-0.028^{**}	-0.032^{**}				
			(0.008)	(0.009)				
$MRPK_{irst-1}(VA)$					0.248^{**}	0.219^{**}	0.225^{**}	0.185^{**}
					(0.016)	(0.016)	(0.023)	(0.023)
$MRPK_{irst-1}$ (VA) * $Median_{rst-1}[\ln \mu_{irst-1}]$					-0.015^{**}	-0.023^{**}	0.014^{+}	-0.006
					(0.005)	(0.007)	(0.008)	(0.010)
$MRPK_{irst-1}$ (VA) * $Fin\ Dep_s$							0.049^{*}	0.067^{**}
							(0.023)	(0.023)
$MRPK_{irst-1}$ (VA) * $Median_{rst}[\ln(\mu_{irst})] * Fin\ Dep_s$							-0.054^{**}	-0.043^{*}
							(0.018)	(0.018)
Plant Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	–	Yes	–	Yes	–	Yes	–
State-sector-year Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	217942	183052	177192	144807	188261	155925	153012	122963

In specifications 1 - 4, MRPK is measured based on gross revenue, and based on value added in specifications 5-8. All specifications control for the logarithm of a plant's age. The variable $Median_{rst-1}[\ln(\mu_{irst-1}]$ is demeaned within 3-digit sectors and measured in standard deviation units. To ensure that the median markup is plausibly exogenous to the individual plant, the sample is restricted to region-sector-years consisting of at least 7 plants. Standard errors, in parentheses, are clustered at the level of 3-digit sectors. $^{+}p < 0.1;^{*} p < 0.05;^{**} p < 0.01$.

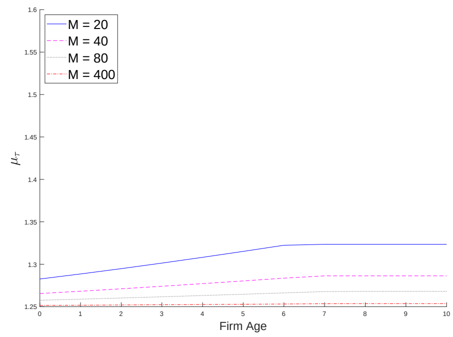
Appendix E Supplementary quantification results

Figure E.1: Impact of competition on markups and capital wedges for different values of σ

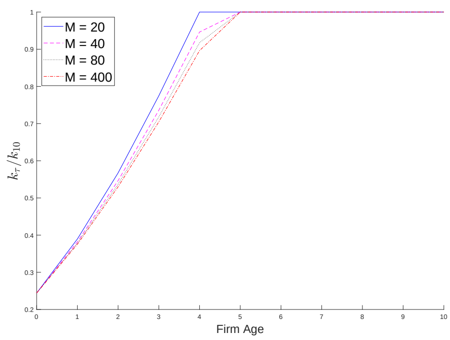
(a) Markup levels when $\lambda = 0.23$, $\sigma = 3$



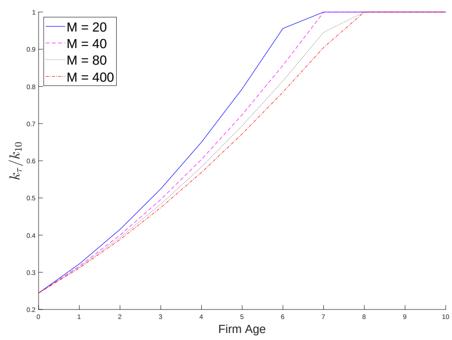
(b) Markup levels when $\lambda = 0.23$, $\sigma = 5$



(c) Capital wedges when $\lambda = 0.23$, $\sigma = 3$



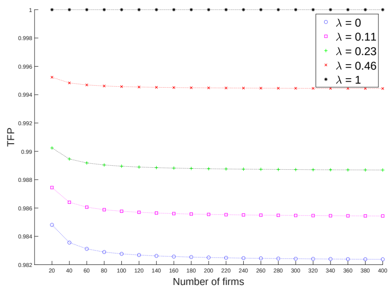
(d) Capital wedges when $\lambda = 0.23$, $\sigma = 5$



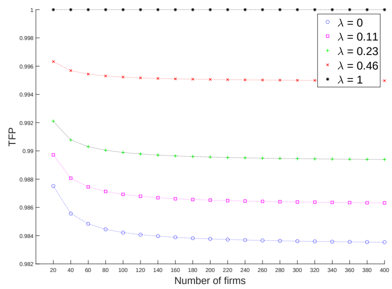
The figure displays the simulation results for an increase in the number of firms M in all sectors. Panels (a) and (b) plot the distribution of markups μ_τ for firms of age τ for $\sigma = 3$ and $\sigma = 4$ respectively, while Panels (c) and (d) plot the distribution of capital wedges for the same respective σ values. A capital wedge is measured as the ratio of the capital level of firms of age τ and age 10, where the latter firm is always unconstrained in these simulations. Since here all firms above age 10 are unconstrained, these firms are omitted from the figure. The value of $\lambda = 0.23$ is the preferred, calibrated value. The σ values are one unit below and above the benchmark value of $\sigma = 4$.

Figure E.2: Impact of competition on aggregate variables for different values of σ

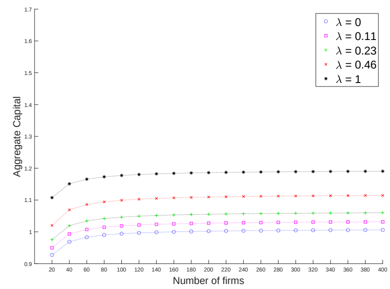
(a) Allocative efficiency when $\sigma = 3$



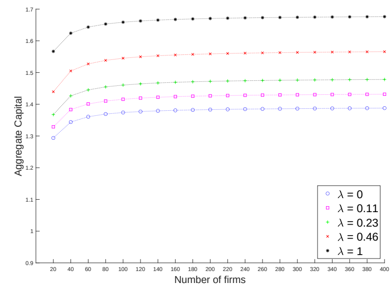
(b) Allocative efficiency when $\sigma = 5$



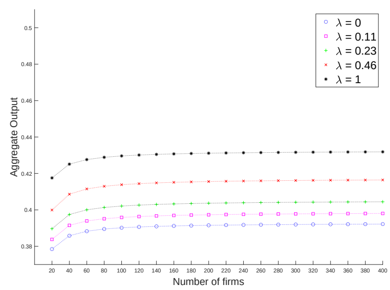
(c) Aggregate capital when $\sigma = 3$



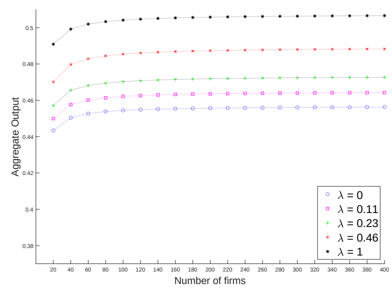
(d) Aggregate capital when $\sigma = 5$



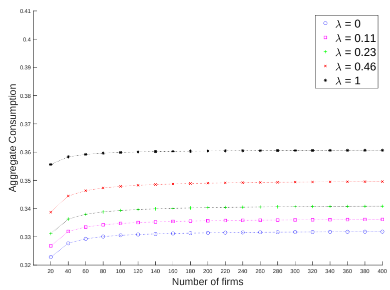
(e) Output when $\sigma = 3$



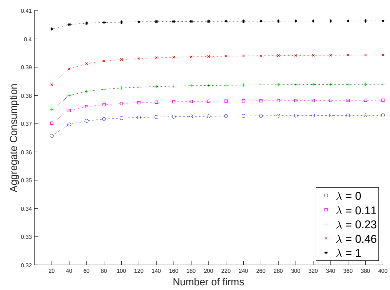
(f) Output when $\sigma = 5$



(g) Consumption when $\sigma = 3$



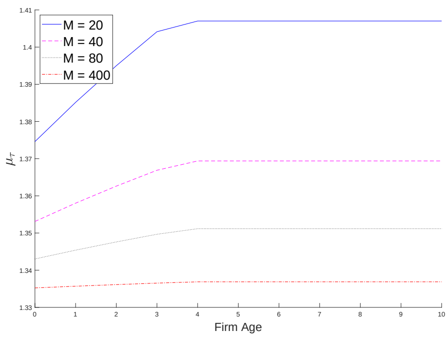
(h) Consumption when $\sigma = 5$



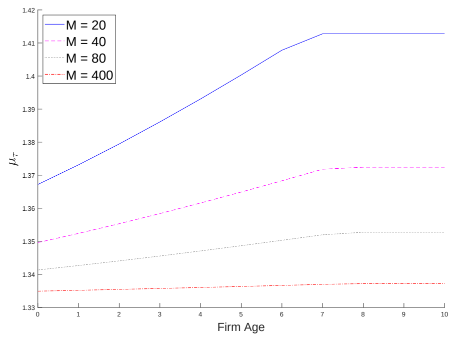
The figure displays the simulation results for an increase in the number of firms M in all sectors, for different values of σ. The preferred, calibrated value for λ is 0.23, and the values 0.11 and 0.46 are (roughly) half and twice that value. No access and perfect access to finance respectively imply $\lambda = 0$ and $\lambda = 1$. Panels (a) and (b) shows the impact of M on TFP; Panels (c) and (d) on aggregate capital across all firms in all sectors; Panels (e) and (f) on output of the final good Q_F; and Panels (g) and (h) on the sum of aggregate worker and firm consumption. The Panels on the left show results for $\sigma = 3$, and those on the right for $\sigma = 5$. These σ values are one unit below and above the benchmark value of $\sigma = 4$.

Figure E.3: Impact of competition on markups and capital wedges for different values of α

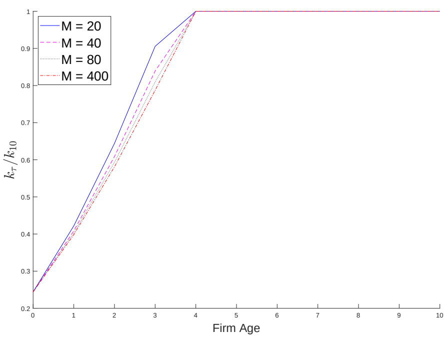
(a) Markup levels when $\lambda = 0.23$, $\alpha = 4/15$



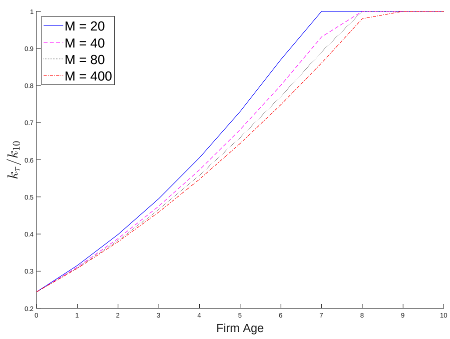
(b) Markup levels when $\lambda = 0.23$, $\alpha = 10/15$



(c) Capital wedges when $\lambda = 0.23$, $\alpha = 4/15$



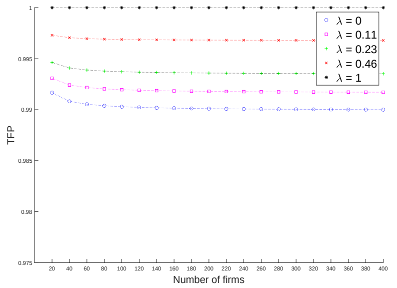
(d) Capital wedges when $\lambda = 0.23$, $\alpha = 10/15$



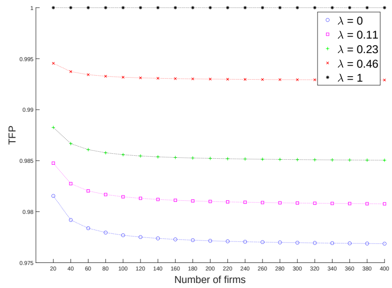
The figure displays the simulation results for an increase in the number of firms M in all sectors. Panels (a) and (b) plot the distribution of markups μ_τ for firms of age τ for $\alpha = 4/15$ and $\alpha = 10/15$ respectively, while Panels (c) and (d) plot the distribution of capital wedges for the same respective α values. A capital wedge is measured as the ratio of the capital level of firms of age τ and age 10, where the latter firm is always unconstrained in these simulations. Since here all firms above age 10 are unconstrained, these firms are omitted from the figure. The value of $\lambda = 0.23$ is the preferred, calibrated value. The α values are $3/15$ below and above the benchmark value of $\alpha = 7/15$.

Figure E.4: Impact of competition on aggregate variables for different values of α

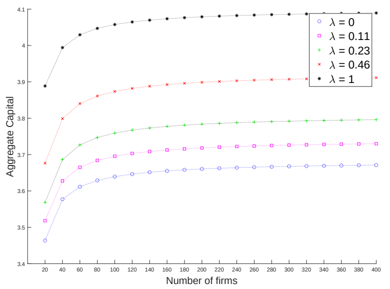
(a) Allocative efficiency when $\alpha = 4/15$



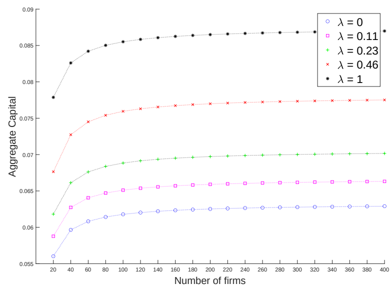
(b) Allocative efficiency when $\alpha = 10/15$



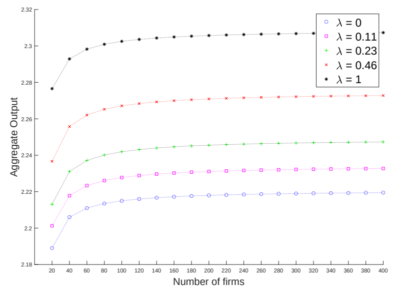
(c) Aggregate capital when $\alpha = 4/15$



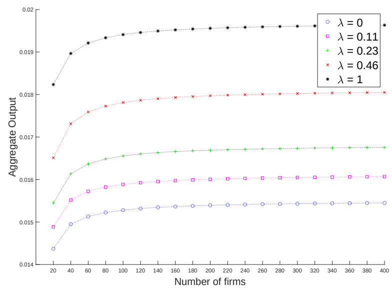
(d) Aggregate capital when $\alpha = 10/15$



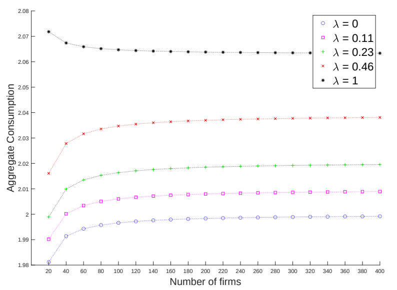
(e) Output when $\alpha = 4/15$



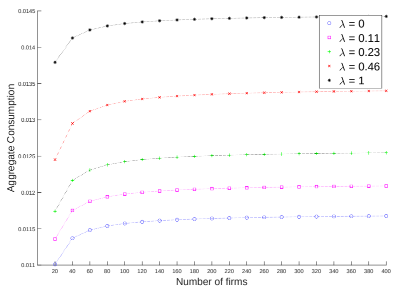
(f) Output when $\alpha = 10/15$



(g) Consumption when $\alpha = 4/15$



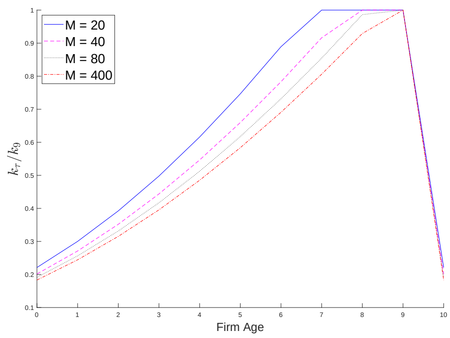
(h) Consumption when $\alpha = 10/15$



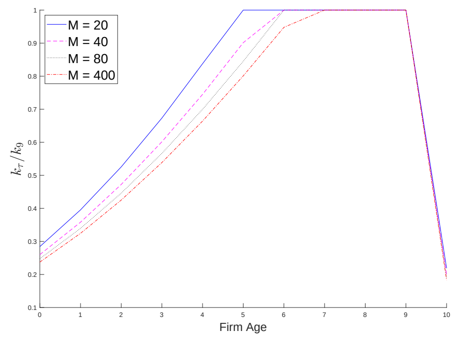
The figure displays the simulation results for an increase in the number of firms M in all sectors, for different values of α. The preferred, calibrated value for λ is 0.23, and the values 0.11 and 0.46 are (roughly) half and twice that value. No access and perfect access to finance respectively imply $\lambda = 0$ and $\lambda = 1$. Panels (a) and (b) shows the impact of M on TFP; Panels (c) and (d) on aggregate capital across all firms in all sectors; Panels (e) and (f) on output of the final good Q_F; and Panels (g) and (h) on the sum of aggregate worker and firm consumption. The Panels on the left show results for $\alpha = 4/15$, and those on the right for $\alpha = 10/15$, which is 3/15 below and above the benchmark value of $\alpha = 7/15$.

Figure E.5: Impact of competition on markups and capital wedges with age-specific η_τ

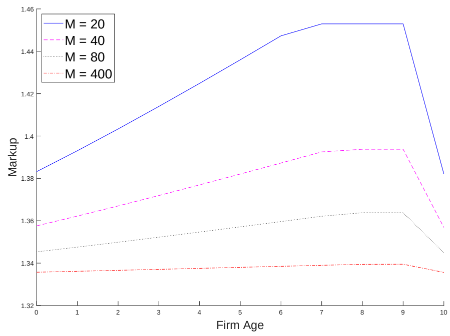
(a) Markup levels when $\lambda = 0$



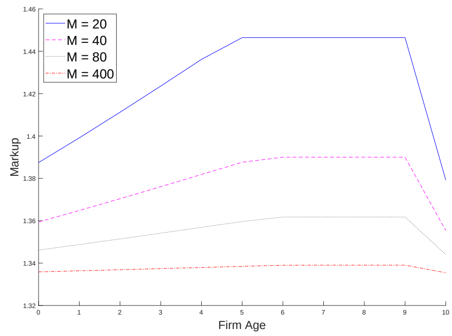
(b) Markup levels when $\lambda = 0.23$



(c) Capital ratios when $\lambda = 0$

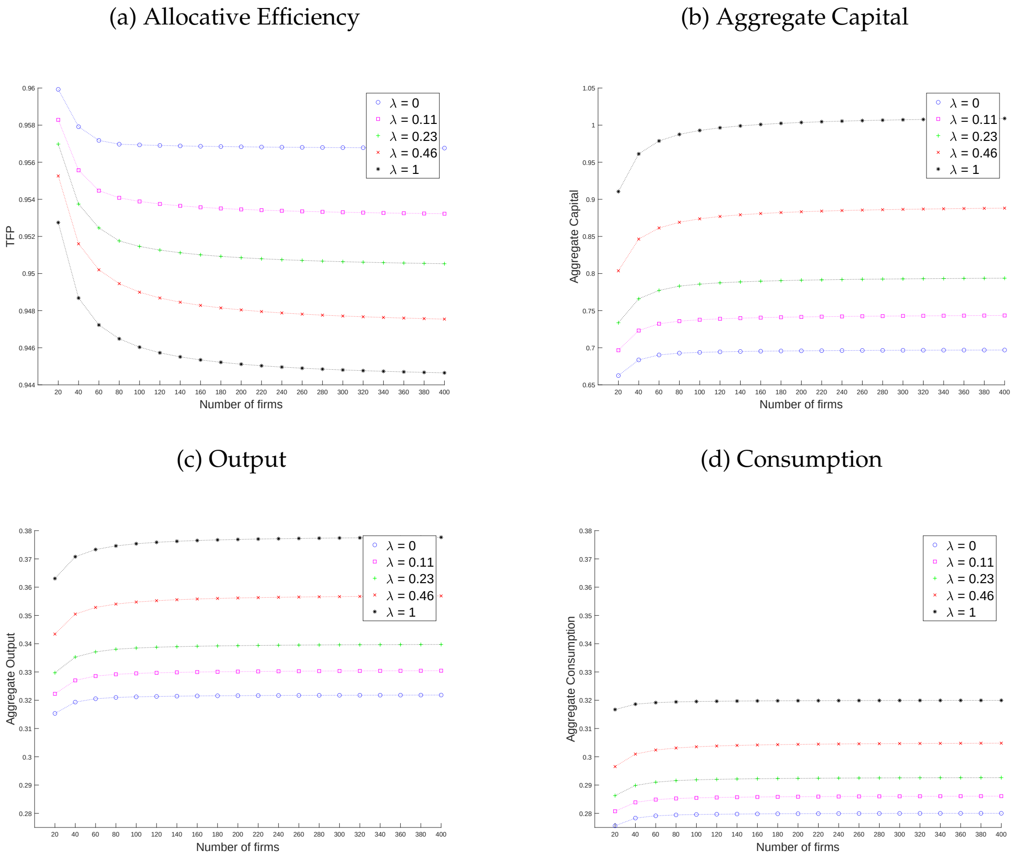


(d) Capital ratios when $\lambda = 0.23$



The figure displays the simulation results for an increase in the number of firms M in all sectors, when firms incorporate their age-specific survival rate η_τ in their discounting. All firms up until age 9 have $\eta_\tau = 1$, and the older firms have $\eta_\tau = 0.9$. They discount future utility at a rate $\beta\eta_\tau$, with $\beta = 0.95$. Panels (a) and (b) plot the distribution of markups μ_τ for firms of age τ for $\lambda = 0$ and $\lambda = 0.23$ respectively, while Panels (c) and (d) plot the distribution of capital ratios for the same respective λ values. The capital ratio is the ratio of the capital level of firms of age τ and age 9, where the latter firm is always unconstrained in the current simulations. All firms aged 10 years and above have the same capital level, which is smaller due to their lower η_τ. This lower capital level also results in lower markups. In these simulations, newborn firms receive all the capital from firms who died. This ensures the difference in log revenue of newborn versus larger firms is at 0.77, which is close to the calibration target of 0.74.

Figure E.6: Impact of competition on aggregate variables with age-specific η_τ



The figure displays the simulation results for an increase in the number of firms M in all sectors, for different values of λ, when firms incorporate their age-specific survival rate η_τ in their discounting. All firms up until age 9 have $\eta_\tau = 1$, and the older firms have $\eta_\tau = 0.9$. They discount future utility at a rate $\beta\eta_\tau$, with $\beta = 0.95$. The preferred, calibrated value for λ is 0.23, and the values 0.11 and 0.46 are (roughly) half and twice that value, respectively. No access and perfect access to finance respectively imply $\lambda = 0$ and $\lambda = 1$. Panel (a) shows the impact of M on TFP_s; Panel (b) on aggregate capital across all firms in all sectors; Panel (c) on output of the final good Q_F; and Panel (d) on the sum of aggregate worker and firm consumption. The results are very similar to those in Figure 3, except for how the impact of competition varies with λ. In the current figure, the decline in TFP is strongest for $\lambda = 1$, this is because markup variation partly offsetts the difference in capital levels between young and old firms, and this markup variation declines with M.