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

Simon Galle *
BI Norwegian Business School

March 2017

Preliminary - Please find the latest version here

Abstract

This paper develops a novel general-equilibrium model of the relationship between competition, financial constraints and misallocation, and tests its implications using Indian plant-level panel data. In the model, steady-state misallocation consists of both variable markups and capital wedges. The variable markups arise from Cournot-type competition, whereas the capital wedges result from the interaction of firm-level productivity volatility with financial constraints. Firms experience random shocks to their productivity and in response to positive productivity shocks they optimally grow their capital stock, subject to financial constraints. Competition plays a dual role in affecting misallocation. On the one hand, both markup levels and markup dispersion tend to fall with competition, which unambiguously improves allocative efficiency in a setting without financial constraints. On the other hand, in a setting with financial constraints, a reduction in markups is associated with slower capital accumulation, as the rate of self-financed investment falls. Thus, the positive impact of competition on steady-state misallocation is reduced by the presence of financial constraints. Empirically, I test and confirm the qualitative predictions of the model with data on Indian manufacturing. First, I exploit natural variation in the level of competition, arising from the pro-competitive impact of India's 1997 dereservation reform on incumbent plants. I show that, in line with the model, this reform lead to a reduction in markup levels and markup dispersion, as well as to a slowdown in the firm-level speed of capital convergence. Finally, I corroborate the external validity of the finding that capital convergence slows down with competition by providing evidence for the full panel of manufacturing plants in India's Annual Survey of Industries, and show that this slowdown is particularly pronounced in sectors with higher levels of financial dependence.

---

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

1 Introduction

Misallocation of resources has recently become a prominent explanation for cross-country differences in economic development. For instance, Hsieh and Klenow (2009) argue that misallocation, arising from the misalignment of marginal products across plants, could account for 40 to 60% of the difference in aggregate output per capita between the United States and India. This finding has sparked a debate on the main driving forces of the pattern in measured misallocation across countries. For instance, measured capital misallocation, which motivates this paper's analysis, can be explained by technological constraints, market imperfections or policy distortions, amongst others.[1] Knowledge on the relative importance of these different underlying mechanisms matters to understand the potential level of macroeconomic efficiency gains from specific policy interventions.

This paper contributes to the above debate by investigating the relationship between competition, financial constraints and misallocation. Existing work explains how in a setting with variable markups, competition reduces misallocation by decreasing dispersion in markups (e.g. Asturias et al. (2017), Peters (2013)).[2] While such a channel is still present in my analysis, I demonstrate that financial constraints introduce a second, negative impact of competition on misallocation. Specifically, I show that competition slows down the capital growth rate of financially constrained firms. Thereby, capital wedges, which result from the difference between the optimal and the actual capital level of a firm, are amplified by competition. The intuition for this result is that firm-level markups fall with the degree of competition, which lowers the rate of internally financed capital growth. I then empirically test and confirm the qualitative predictions of the model with data on the Indian manufacturing sector.

In the model, capital misallocation arises due to the interaction of productivity volatility and financial constraints. Productivity volatility in this context means that firms experience random shocks to their idiosyncratic levels of productivity. After a positive productivity shock, a firm will optimally choose to grow its capital stock, but the financial constraint will limit its ability to do so. A financially constrained firm will therefore rely on internally financed capital growth, which will imply that the firm's capital growth is a function of its markup. Since the firm's capital growth rate depends on its markup, its speed of convergence to its optimal level of capital will also depend on the markup. Increased competition, by reducing a firm's

[1] Roughly speaking, measured capital misallocation is a function of the dispersion in marginal revenue products of capital (MPRK). As such it is a salient component of aggregate misallocation. Asker et al. (2014) propose a model where such dispersion in MRPK is explained by adjustment costs in capital, which is a form of technological constraints. In this case, the dispersion in MRPK is the consequence of first-best optimization, and does not constitute a *mis*allocation of capital. In other settings measured dispersion in MRPK arises from market imperfections or policy distortions. In the models presented by Midrigan and Xu (2014) and Moll (2014), capital misallocation is driven by firm-level collateral constraints, which arise from imperfect financial markets. Restuccia and Rogerson (2008), in their seminal contribution to the misallocation literature, model misallocation as the result of firm-level variation in taxes or subsidies, which results from e.g. a non-competitive banking sector varying its interest rates for noneconomic reasons. Restuccia and Rogerson (2013) provide a broader survey of the misallocation literature, while Buera et al. (2015) survey the literature on the macro-economic impact of financial constraints.

[2] In an earlier contribution, Epifani and Gancia (2011) demonstrate that trade liberalization can have ambiguous effects on markup misallocation. In their setting, misallocation arises from differences in the degree of competition across industries, whereas I focus on varying the degree of competition within a single industry.

markup, will then negatively affect its speed of capital convergence in response to a positive productivity shock. This way, capital wedges are amplified by competition.

A related channel through which competition can negatively affect capital misallocation applies to young plants. I present this channel in a version of the model where there is no productivity volatility but instead there is birth and death of firms. If newborn firms are undercapitalized and therefore financially constrained, these firms will also rely on internal financing while converging to their optimal level of capital. This implies that competition again reduces the speed of capital convergence and thereby amplifies capital wedges.[3]

I then test the predictions of the model in the context of the Indian manufacturing sector. I first test the main mechanism of the model, namely that firm-level speed of capital convergence decreases with competition. This prediction can be tested at two levels: for firms in general and for young plants in particular. For firms in general, I test whether, after a firm deviates from its optimal marginal revenue product of capital (MRPK), it converges back faster to its optimal MRPK in a setting with less competition. Then, based on the model's prediction for undercapitalized young plants, I check if the capital growth rate of young plants is faster in settings where competition is less intense. These two empirical tests are complementary. The first test is closely linked to the structure of the model with productivity volatility as it focuses directly on plant-level MRPK, where, inspired by Asker et al. (2014), plant-level deviations in MRPK serve as a proxy for the plant-level capital wedges. The second test, which focuses on young plants, has the advantage that capital growth is a reduced-form object in the data, and therefore relies on fewer assumptions for its measurement. The fact that both tests empirically validate the model predictions, therefore provides robust support for the model.

A second set of tests leverages heterogeneity in firms' financial dependence, where capital convergence of firms in sectors with higher financial dependence exhibits a stronger sensitivity to the degree of competition. To test this prediction, I augment the baseline tests with an interaction term of the competition measure with Rajan and Zingales (1998) measures of sector-level financial dependence. The data again support the predictions, both for the test on MRPK convergence, and for the test on capital growth for young firms.

These two sets of predictions rely on a measure for competition that is arguably exogenous from the individual plant's point of view, namely the median markup measured at the state-sector-year level. The advantage of this approach is that I can test the theory on a large set of Indian manufacturing plants, while a potential limitation is that the underlying structural drivers of the variation in the levels of competition remain unexamined. To address this concern, I also exploit natural variation in the degree of competition arising from India's 1997 dereservation reform. After demonstrating the pro-competitive impact of the dereservation reform on incumbent plants, I now test whether after the reform, MRPK convergence and capital growth of young plants is slower.[4] The data again confirm the two predictions of the model.

[3] I provide evidence that productivity volatility and the birth of newborn firms are both contributing to capital misallocation in the Indian manufacturing sector.

[4] The dereservation reform gradually removed previously existing investment ceilings on a set of "reserved" products (García-Santana and Pijoan-Mas, 2014; Martin et al., 2014; Tewari and Wilde, 2016). Hence, the direct effect of dereservation is to allow incumbent firms to increase their capital stock. However, the reform also leads

The theory builds on Midrigan and Xu (2014), who examine comparative statics for steady-state capital misallocation in a setting of imperfect competition. The main focus of Midrigan and Xu (2014) is on quantifying the relative importance of barriers to entry for new firms versus collateral constraints for incumbent firms in shaping misallocation. My focus on the comparative statics for competition is therefore complementary to their analysis. Moreover, I analytically derive general theoretical results, whereas Midrigan and Xu (2014) rely on simulation-based methods. Interestingly, the steady state in my model is determined by a system of non-linear equations, which does not allow for a closed-form solution to the comparative statics exercise. Instead, I derive such a solution by exploiting the logical properties of the steady state equilibrium.

Moll (2014) and Itskhoki and Moll (2015) also analyze capital misallocation analytically. However, they do so in a setting of perfect competition, whereas I study the impact of varying competition on misallocation. This difference in market structure not only makes our analyses complementary, it also modifies the theoretical solution strategy because a closed-form solution is available under perfect competition, but not under imperfect competition.

By examining the potential downsides of intensified competition, this paper complements papers that emphasize the beneficial impacts of competition on misallocation.[5] For instance, Peters (2013) argues that increased competition diminishes misallocation, as it reduces the dispersion in the distribution of markups. A second, well-established, beneficial impact of competition consists in reallocating labor from low productivity to high productivity firms. Here, Melitz (2003) studies the role of trade liberalization in improving the allocative efficiency of labor, and Akcigit et al. (2014) analyze constraints to such reallocation through competition in a Schumpeterian growth model with firm-level limits to delegation.

The analysis by Akcigit et al. (2014) is motivated by the stylized fact on firm-stagnation in India (Hsieh and Klenow, 2014). Such slow growth of firms is part of the broader lack of reallocation and persistent level of misallocation in the Indian manufacturing sector, as analyzed by Bollard et al. (2013). In this paper, I aim to contribute to our understanding of this high and persistent level of misallocation in the Indian manufacturing sector. As indicated above, the adverse effect of competition depends, amongst others, on the degree of productivity volatility and the entry-rate of newborn firms in a context of financial constraints. Existing stylized facts strongly suggest that both productivity volatility and entry of newborn firms, two possible sources of misallocation in a setting with financial constraints, are potentially important for misallocation in Indian manufacturing. First, for productivity volatility, Asker et al. (2014) demonstrate that there is a strong correlation between productivity volatility and their measure of capital misallocation in the case of India. Second, Bollard et al. (2013) document high entry-rates of new firms in Indian manufacturing. Third, Banerjee and Duflo (2014) estimate

to intensified competition, e.g. through larger firms starting to produce the previously reserved products, and this competitive channel empirically dominates in my analysis of capital convergence.

[5] In the innovation literature, it is well-established that increasing competition can have both positive and negative impacts on aggregate output (see e.g. Aghion et al. (2005, 2013); Gilbert (2006)). Also in the empirical micro-development literature there is work that studies the downsides of competition (Macchiavello and Morjaria, 2015). However, in the misallocation literature, the downsides of misallocation have been understudied.

severe credit constraints for large Indian firms, which is consistent with the descriptive evidence on financial constraints from the World Bank Enterprise Surveys (Kuntchev et al., 2014). Together, these three stylized facts on productivity volatility, arrival rate of newborn firms, and financial constraints, indicate the relevance of this paper for understanding misallocation in Indian manufacturing.

2 Theory

2.1 Setup of the economy

Agents The economy has two types of agents: workers and firm owners. The measure L of workers supplies labor inelastically, and each worker is hired at a wage w_t, where t indicates the time period. A worker's consumption c_{lt} is hand-to-mouth.

There is an exogenous, finite set M of firm-owners.[6] Firm-owner i has the following intertemporal preferences at time s:

$$U_{it} = \sum_{t=s}^{\infty} \beta^{t-s} d_{it}$$

Where β is the discount factor and d_{it} is firm-owner consumption.[7]

Production of varieties Each firm produces a variety i with a Cobb-Douglas production function, using capital k_{it} and labor l_{it} as inputs:

$$y_{it} = a_{it} k_{it}^{\alpha} l_{it}^{1-\alpha} \tag{1}$$

Productivity a_{it} follows a stochastic process over the state space $a_{it} \in \{a_L, a_H\}$, where $a_L < a_H$.[8] Firm-level productivity volatility, arising from this stochastic path of a_{it}, will be central in the analysis of steady-state firm-dynamics in section 2.4. Importantly, capital is a dynamic input, subject to the equation of motion:

$$k_{it+1} = x_{it} + (1-\delta) k_{it}$$

with investment x_{it} taking place, and being financed at the end of period t. The decision about labor l_{it+1} is also made in period t, i.e. at the same time the decision on k_{it+1} is made, but labor l_{it+1} is only paid at the end of period $t+1$.[9]

[6] The main comparative statics within the model will be on M, as modifying the degree of competition. Having M as exogenous simplifies the analytical solution of the model. In a simulation-based methodology, as employed by Midrigan and Xu (2014), one can endogenize the degree of competition.

[7] The simplifying assumption of linear firm-owner preferences will prove useful in the analytical derivation of a global solution for the firm-level path of capital.

[8] Increasing the dimensionality of the state space would add substantial complexity to the comparative-statics exercise, without yielding additional economic insight. I follow Midrigan and Xu (2014) by assuming that in period t, the firm is informed about the distribution of a_{it+1}.

[9] The assumption of labor and capital being decided simultaneously, will simplify the optimization problem.

Demand Investment x_{it}, workers' consumption c_{lt} and firm-owner consumption d_{it} all consist of shares of the final good Q_t, which is composed of varieties q_{it}:

$$Q_t \equiv M^{1-\frac{1}{\eta}} \left[\sum_{i=1}^{M} q_{it}^{\eta} \right]^{\frac{1}{\eta}} \tag{2}$$

where $M^{1-\frac{1}{\eta}}$ eliminates taste-for-variety (Blanchard and Kiyotaki, 1987).[10] [11] This expression for the composite good implies that firms face the following demand function q_{it}:

$$q_{it} = \left(\frac{p_{it}}{P_t} \right)^{-\frac{1}{1-\eta}} M^{-\frac{1}{\eta}} \left[\sum_{i=1}^{M} q_{it}^{\eta} \right]^{\frac{1}{\eta}} \tag{3}$$

where p_{it} is the price of variety i and P_t is the price of the final good:

$$P_t^{-\frac{\eta}{1-\eta}} \equiv \frac{1}{M} \sum_{i=1}^{M} p_{it}^{-\frac{\eta}{1-\eta}} \tag{4}$$

Financial constraint The above implies that firms face the following period-by-period budget constraint, where z_{it} is wealth at the end of period t: $z_{it} \equiv p_{it}y_{it} - w_t l_{it} + P_t(1-\delta)k_{it}$.

$$k_{it+1} + d_{it} \leq \frac{z_{it}}{P_t} \tag{5}$$

The financial constraint implies that consumption d_{it} cannot be negative:

$$d_{it} \geq 0 \tag{6}$$

2.2 Firm's problem

Market structure and firm problem I follow Atkeson and Burstein (2008) by assuming that each period, firms play a one-period game of quantity competition.[12] Specifically, each firm i sets a quantity y_{it+1} for sale, conditional on the quantities chosen by the other firms in the economy. As discussed in the previous subsection, firms make decisions about l_{it+1}, k_{it+1} in period t, knowing a_{it+1} and given the budget constraint $P_t(k_{it+1} + d_{it}) \leq z_{it}$. Therefore, any firm i's optimal decisions are $k_{it+1}\left(a_{it+1}, z_{it}, \mathbf{y_{-it+1}}\right)$, $l_{it+1}\left(a_{it+1}, z_{it}, \mathbf{y_{-it+1}}\right)$, where (a_{it+1}, z_{it}) characterizes the state for firm i and $\mathbf{y_{-it+1}}$ is the vector of decisions on y_{jt+1} for all $j \neq i$. Through the production function (1), the choice of k_{it+1}, l_{it+1} determines y_{it+1} and thereby $p_{it+1}(y_{it+1}, \mathbf{y_{-it+1}})$ as firms incorporate the demand function (3) into their optimiza-

[10] This expression for the final good is employed by Jaimovich (2007) in a setting with variable markups, and it allows to restrict attention to the competitive effects of varying M, and ignore the taste-for-variety effects. Bénassy (1996) generalizes the idea of de-linking consumption-side taste-for-variety and firm-level market power.

[11] There is one sector, and Q_t is the composite good of that sector. Note that it should be straightforward to extend this to a multi-sector case when preferences are Cobb-Douglas across sectors, as expenditure shares are constant across sectors in that case.

[12] I will assume that strategic interaction of firms is only within-period.

tion. As such, this setting entails the following intertemporal problem for the firm, where $\pi_{it}(k_{it}, l_{it}, \mathbf{y_{-it}}) \equiv p_{it}(y_{it}, \mathbf{y_{-it}})y_{it} - w_t l_{it}$:

$$
\begin{aligned}
\max_{d_{it}, k_{it+1}, l_{it+1}} \mathcal{L} = & \sum_{t=s}^{\infty} E_s \left[\beta^{t-s} d_{it}\right] + \\
& \sum_{t=s}^{\infty} E_s \left[\lambda_{it} \left(\frac{\pi_{it}(k_{it}, l_{it}, \mathbf{y_{-it}})}{P_t} + (1-\delta)k_{it} - k_{it+1} - d_{it}\right) + \Phi_{it}(d_{it})\right]
\end{aligned} \tag{7}
$$

Since each firm's decision on y_{it+1} depends on $(a_{it+1}, z_{it}, \mathbf{y_{-it+1}})$, $\mathbf{y_{it+1}}$ will be determined by $F(a(t+1), z(t))$, the joint distribution of a_{it+1} and z_{it}, and by the conditions in the labor and goods market implied by M, L.

$$
\begin{aligned}
& k_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right) \\
& l_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right)
\end{aligned} \tag{8}
$$

The optimal choices in (8) determine $p_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right)$, and given the firm's marginal cost thereby also determine the markup μ_{it+1}

$$
\mu_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right) = \frac{\varepsilon_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right) - 1}{\varepsilon_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right)} \tag{9}
$$

where the demand elasticity ε_{it} is:

$$
\varepsilon_{it+1}\left(a_{it+1}, z_{it}, F(a(t+1), z(t)), M, L\right) = -\frac{1}{1-\eta} + \left(\frac{\eta}{1-\eta}\right) \frac{y_{it+1}^{\eta}}{\sum_i y_{it+1}^{\eta}} \tag{10}
$$

Labor optimization The first-order condition for labor is standard:

$$
E_s \left[\frac{\partial[\pi_{it}(k_{it}, l_{it}, F(a(t+1), z(t)), M, L)/P_t]}{\partial l_{it}}\right] = 0 \tag{11}
$$

Intertemporal optimization Now I derive the first-order conditions for the dynamic part of the problem. Start with the first-order condition for d_{it}.

$$
\frac{\partial \mathcal{L}}{\partial d_{it}} = \beta^{t-s} + E_s[-\lambda_{it} + \Phi_{it}] = 0 \tag{12}
$$

Next, the first-order condition for k_{it+1} implies:

$$
E_s\left[\lambda_{it}\right] = E_s\left[\lambda_{it+1}\left((1-\delta) + \frac{\partial[\pi_{it+1}(k_{it+1}, l_{it+1}, F(a(t+1), z(t)), M, L)/P_{t+1}]}{\partial k_{it+1}}\right)\right] \tag{13}
$$

2.2.1 Decision rules for capital and consumption

Capital and consumption The combination of (12) and (13) allows me to find the decision rules for d_{it}, k_{it+1}. Taking the perspective of period $s = t$, there are then two cases, either $\Phi_{it} > 0$ or $\Phi_{it} = 0$.

- **Case 1** When $\Phi_{it} = 0$, then k_{it+1} is optimally set such that:[13]

$$1 = E_t \left[\lambda_{it+1} \left((1-\delta) + \frac{\partial[\pi_{it+1}(k_{it+1}, l_{it})/P_{t+1}]}{\partial k_{it+1}} \right) \right] \tag{14}$$

  And consumption $d_{it} = \frac{\pi_{it}(k_{it}, l_{it})}{P_t} - x_{it}$.

- **Case 2** When $\Phi_{it} > 0$, then $d_{it} = 0$ and the path of capital is determined by the budget constraint: $k_{it+1} = \frac{\pi_{it}(k_{it}, l_{it})}{P_t} + (1-\delta)k_{it}$.

Output and markup The above decision rules also imply an output decision for both cases.

- **Case 1** When $\Phi_{it} = 0$, then firms in period t solve the following system of decision rules regarding period $t+1$:

$$E_t \left[\lambda_{it+1} \frac{\partial[\pi_{it+1}(k_{it+1}, l_{it})/P_{t+1}]}{\partial k_{it+1}} \right] = 1 - E_t \left[\lambda_{it+1}(1-\delta) \right]$$
$$\frac{\partial[\pi_{it+1}(k_{it+1}, l_{it+1})/P_{t+1}]}{\partial l_{it+1}} = 0$$

- **Case 2**: When $\Phi_{it} > 0$, then the optimal labor choice l_{it+1} is chosen conditional on $k_{it+1} = \frac{\pi_{it}(k_{it}, l_{it})}{P_t} + (1-\delta)k_{it}$.

Given the decision on k_{it+1}, l_{it+1}, the output y_{it+1} is determined due to the production function (1). Then, given (3), this determines the price p_{it+1} of the firm. This pricing decision simultaneously implies a decision on the markup in (9), given the firm's marginal cost.

2.3 Steady state equilibrium

An *equilibrium* consists of a set of prices P_t, w_t, p_{it}, a set of consumption $d_{it}(a_{it+1}, z_{it}, F(a(t+1), z(t)))$, capital $k_{it+1}(a_{it+1}, z_{it}, F(a(t+1), z(t)))$ and labor $l_{it}(a_{it}, z_{it-1}, F(a(t), z(t-1)))$ decisions by firm-owners and consumption by workers $\frac{w_t}{P_t}L$ that satisfy

- the labor market clearing condition

$$L = \sum_{i=1}^{M} l_{it} \tag{15}$$

- the goods market clearing condition

[13] When $E_t[\Phi_{it+1}] = 0$, then $E_t[\lambda_{it+1}] = \beta$, and therefore (14) simplifies to $\frac{\partial[\pi_{it+1}(k_{it+1}, l_{it})/P_{t+1}]}{\partial k_{it+1}} = \frac{1}{\beta} + \delta - 1$.

$$Q_t = \sum_{i=1}^{M}(x_{it} + d_{it}) + \int_{l\in L} c_{lt} dl \tag{16}$$

- the optimality conditions (11), (13) for each firm i, conditional on the choices of l_{jt}, k_{jt} of all firms $j \neq i$.
- market-clearing for each variety i: $y_{it} = q_{it}$, satisfying (3)
- the equalized budget constraint $P_t(k_{it+1} + d_{it}) = z_{it}$, and the financial constraint $d_{it} \geq 0$.

To solve this equilibrium, I can pick as numeraire $w_t = 1$, and P_t is a function of the individual prices as in (4). Next, y_{it} is determined by k_{it}, l_{it}, a_{it}, where a_{it} is exogenous. Satisfying (3) implies that p_{it} is given by choice of y_{it}. Finally, l_{it}, k_{it}, d_{it} are determined by (11), (13) and the budget constraint (5), as explained in section 2.2.1. Since there are M firms, this then is a system of Mx3 equations with Mx3 unknowns.

A *steady state equilibrium* is an equilibrium that satisfies for all t[14] :

$$\begin{aligned} K_t &= K, \\ \frac{P_t}{w_t} &= \frac{P}{w}, \\ F(a(t+1), z(t)) &= F(a', z) \end{aligned} \tag{17}$$

A first implication of this definition of the steady state, is that $H(a(t), k(t)) = H(a, k)$, i.e. the joint distribution of productivities and capital will be stable.[15] The reason is that capital choice is determined by $F(a(t+1), z(t))$: $k_{it+1}(a_{it+1}, z_{it}, F(a(t+1), z(t)))$. A second implication is that aggregate output will be stable as well: $Q_t = Q$.

2.4 Analysis of the steady state

Section 2.3 implies that in steady state each firm's decisions depend on $F(a+1, z)$. Here, the wealth distribution is endogenous, whereas the distribution of productivities is exogenously determined. Since the distribution of wealth is a function of $H(a, k)$, I focus on examining this joint distribution of productivities and capital in steady state. To this end, I will start by characterizing the firm's decision rules for capital and labor in steady state.

2.4.1 Labor and capital decisions in steady state

It will be convenient to characterize the solution to the firm's optimization problem by taking the perspective of the cost-minimization problem given the optimal markup characterized in

[14] Moll (2014) employs a similar definition of a steady state equilibrium.

[15] The assumptions on the productivity volatility process, described in Appendix C, are such that the productivity volatility process allows for a stable $H(a, k)$.

(9).[16] As such, the cost-minimization problem implies the following optimal labor demand in steady state:

$$l_{it} = \left(\frac{(1-\alpha)}{\mu_{it}} \frac{P}{w} \left(\frac{Q}{M} \right)^{1-\eta} a_{it}^{\eta} k_{it}^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} \tag{18}$$

For the capital choice, as is clear from section 2.2.1, there are two cases: either $\Phi_{it} = 0$, or $\Phi_{it} > 0$.

Unconstrained firms First consider the case where a firm has $\Phi_{it} = 0$. In that case, the optimality condition in (13), together with (18) implies that

$$k_{it}^{*} = \mu_{it}^{\frac{1}{\eta-1}} a_{it}^{\frac{\eta}{1-\eta}} \frac{Q}{M} \left(\frac{P(1-\alpha)}{w} \right)^{\frac{\eta-\alpha\eta}{1-\eta}} \left(\frac{\alpha}{r_{it}} \right)^{\frac{1+\alpha\eta-\eta}{1-\eta}} \tag{19}$$

where $r_{it} \equiv \left(\frac{1}{\beta} + \delta - 1 \right) - \xi_{it}$.[17]

Constrained firms When the financial constraint binds, i.e. $\Phi_{it} > 0$. Capital grows according to the budget constraint. Specifically, I show in appendix B.2 that:

$$k_{it+1} = (1-\delta)k_{it} + [\mu_{it} - (1-\alpha)] \frac{w_t}{P_t(1-\alpha)} l_{it} \tag{20}$$

2.4.2 Distribution and dynamics for firm-level capital

Given the expressions for k_{it}^*, and the path for capital of constrained firms in (20), I now characterize $H(a,k)$. First, consider the firms with $a_{it} = a_L$. In steady state, these firms cannot have $\Phi_{it} > 0$[18], and therefore these firms have $k_{it} = k_L^*$, the optimal level of k_{it} for low productivity firms. Note that $k_{it}(a_L) > k_L^*$ violates the firm's optimality conditions, as firms consume any capital in excess of k_L^*, and thereby satisfy the decision rule for capital in equation (19).

Second, there are the firms with $a_{it} = a_H$. For these firms, either $\Phi_{it} = 0$, or $\Phi_{it} > 0$. When $\Phi_{it} = 0$, then these firms have $k_{it} = k_H^*$. When $\Phi_{it} > 0$, then $k_{it} = G_\tau k_L^*$, where $\tau = t - s$,

$$G_\tau \equiv \Pi_{r=s}^{s+\tau}(1+g_r) \tag{21}$$

and

$$g_r \equiv \frac{k_{r+1}}{k_r} - 1; \; s \equiv \max r \text{ s.t. } a_{ir+1} = a_H \& a_{ir} = a_L$$

Here, k_{r+1} is determined by (20), for any firm i with capital level k_r. In words, k_{it} is determined by the cumulative capital growth G_τ since the firm's most recent positive productivity

[16] Jaimovich (2007) also employs the cost-minimization approach to characterize the solution to the firm problem, and as such, the optimality conditions are closely related to the ones found in that paper.

[17] When $E_t[\Phi_{t+1}] = 0$, then $\xi_{it} = 0$, otherwise $\xi_{it} > 0$.

[18] Suppose this is not the case and there is at least one firm with $a_{it} = a_L \& \Phi_{it} > 0$. Then for all firms i with $a_{it} = a_L \& \Phi_{it} > 0$, $k_{it+1}(a_{it+1}, z_{it}, F(a+1,z)) > k_{it}$. Since these firms will grow their capital, for them $k_{it+1} > k_{it}$. Given that firms are infinitely lived, this would then violates the property of the steady state that $F(a(t+1), k(t)) = F(a', z)$.

shock.

Capital of unconstrained firms Following (19), the optimal values for capital k_L^*, k_H^* are:

$$\begin{aligned} k_L^* &= \left(\frac{a_L^\eta}{\mu_L}\right)^{\frac{1}{1-\eta}} \left(\frac{\alpha}{r_L}\right)^{\frac{1+\alpha\eta-\eta}{1-\eta}} \left(\frac{P(1-\alpha)}{w}\right)^{\frac{\eta-\alpha\eta}{1-\eta}} \frac{Q}{M} \\ k_H^* &= \left(\frac{a_H^\eta}{\mu_H}\right)^{\frac{1}{1-\eta}} \left(\frac{\alpha}{r_H}\right)^{\frac{1+\alpha\eta-\eta}{1-\eta}} \left(\frac{P(1-\alpha)}{w}\right)^{\frac{\eta-\alpha\eta}{1-\eta}} \frac{Q}{M} \end{aligned} \tag{22}$$

Where μ_L, μ_H, characterized further in section 2.4.3, are the optimal level of markups for the respective firms. Furthermore, $r_H = \frac{1}{\beta} + \delta - 1$ since $E_t[\Phi_{it}] = 0$ for all firms with $a_{it} = a_H$ and $\Phi_{it} = 0$. Next, r_L is the value for r_{it} for all firms with $a_{it} = a_L$. Since for firms with $a_{it} = a_H$, the level of capital depends on G_τ, the value of Φ_{it} is also determined by τ, i.e. the number of periods since the most recent productivity shock. The above entails that the following lemma holds.

Lemma 1. *Steady state $H(a,k)$ is determined by:*

- *if $a_{it} = a_L$, then $k_{it} = k_L^*$*
- *if $a_{it} = a_H$ then $\forall i$ with $\tau = t - s$, where $s = \max r$ s.t. $a_{ir+1} = a_H \& a_{ir} = a_L$:*
  - *if $\Phi_\tau = 0$, then $k_{i\tau} = k_H^*$*
  - *if $\Phi_\tau > 0$, then $k_{i\tau} = G_\tau k_L^*$*

2.4.3 Distribution of markups

Now, I characterize the distribution of markups. First, the markups for the unconstrained firms follow directly from (9), (10) and Lemma 1.

$$\begin{aligned} \mu_L(a_L, k_L^*, H(a,k), M) &\equiv \frac{1 - M^{\eta-1}\eta\frac{(a_L(k_L^*)^\alpha(l_L^*)^{1-\alpha})^\eta}{Q^\eta}}{\eta\left(1 - M^{\eta-1}\frac{(a_L(k_L^*)^\alpha(l_L^*)^{1-\alpha})^\eta}{Q^\eta}\right)} \\ \mu_H(a_H, k_H^*, H(a,k), M) &\equiv \frac{1 - M^{\eta-1}\eta\frac{(a_H(k_H^*)^\alpha(l_H^*)^{1-\alpha})^\eta}{Q^\eta}}{\eta\left(1 - M^{\eta-1}\frac{(a_H(k_H^*)^\alpha(l_H^*)^{1-\alpha})^\eta}{Q^\eta}\right)} \end{aligned} \tag{23}$$

Constrained firms For constrained firms, we know that $k_{it} = G_\tau k_L^*$, and the markup for these firms can be written as:

$$\mu_\tau(a_H, G_\tau k_L^*, H(a,k), M) \equiv \frac{1 - M^{\eta-1}\eta\frac{(y(a_H, G_\tau k_L^*), F(a,k))^\eta}{Q^\eta}}{\eta\left(1 - M^{\eta-1}\frac{(y(a_H, G_\tau k_L^*), F(a,k)))^\eta}{Q^\eta}\right)} \tag{24}$$

Together (23), (24), characterize the distribution of markups.

2.4.4 Capital wedges

Next, I analyze the capital wedges ω_{it}, which will be important in the analysis of aggregate TFP. The capital wedges are implicitly defined in the following way:

$$k_{it} = \left(\frac{a_{it}^{\eta}}{\mu_{it}}\right)^{\frac{1}{1-\eta}} \left(\frac{\alpha}{\omega_{it}}\right)^{\frac{1+\alpha\eta-\eta}{1-\eta}} \left(\frac{P(1-\alpha)}{w}\right)^{\frac{\eta-\alpha\eta}{1-\eta}} \frac{Q}{M} \tag{25}$$

where $\omega_{it} = r_L, r_H$ for unconstrained firms with productivities a_L, a_H respectively, and $\omega_{it} > r_H$ for constrained firms. For these constrained firms, I combine equations (22) and (25), to express the capital wedge for any period τ:[19]

$$\omega_\tau = G_\tau^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_H^{\eta}}{a_L^{\eta}} \frac{\mu_L}{\mu_\tau}\right]^{\frac{1}{1+\alpha\eta-\eta}} r_L \tag{26}$$

Note that: $\max_t \omega_\tau = \omega_1 = G_1^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_H^{\eta}}{a_L^{\eta}} \frac{\mu_L}{\mu_1}\right]^{\frac{1}{1+\alpha\eta-\eta}} r_L$. Hence the distribution of ω_τ for firms with $a_{it} = a_H$, has a range $[r_H, \omega_1]$.

Lemma 2. *In steady state, the distribution of capital wedges is:*

- *For firms with* $a_{it} = a_L$, $\omega_{it} = r_L$
- *For firms with* $a_{it} = a_H$*:*
  - *When* $\Phi_\tau = 0$, $\omega_{it} = r_H$
  - *When* $\Phi_\tau > 0$, $\omega_\tau(G_\tau, \mu_\tau) = G_\tau^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_H^{\eta}}{a_L^{\eta}} \frac{\mu_L}{\mu_\tau}\right]^{\frac{1}{1+\alpha\eta-\eta}} r_L$

2.4.5 Aggregates for output, capital and TFP

Aggregate output In appendix A.1, I show that

$$Q = TFPK^{\alpha}L^{1-\alpha} \tag{27}$$

where TFP is aggregate productivity and K is aggregate capital.

TFP I now characterize TFP. In appendix A.1, I derive equation (46), which is the explicit function for TFP. It is clear from that equation, that TFP is a function of the joint distribution of productivities, markups and capital wedges ω_{it}. Since the capital wedges are a function of a_{it}, k_{it}, I can use Lemma 1 and equations (23),(24), to characterize TFP as:

$$TFP = F_{TFP}(H(a,k), M) \tag{28}$$

[19] The expression is found after simplifying $\omega_{it} = \alpha(G_{it,s}k_L^*)^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_{it}^{\eta}}{\mu_{it}} \left(\frac{Q_t}{M}\right)^{1-\eta} \left(\frac{P_t(1-\alpha)}{w_t}\right)^{\eta-\eta\alpha}\right]^{\frac{1}{1+\alpha\eta-\eta}}$

Aggregate capital Given Lemma 1, aggregate capital $K_t = \sum_{i=1}^{M} k_{it}$ can in steady state be expressed as:

$$K = M\left[Prob(a_{it} = a_L)k_L^* + \sum_{\tau=1}^{\infty} Prob(a_{it} = a_H \& s = t-\tau)G_\tau k_L^*\right]$$

After substituting in the value for k_L^*, and using $Q = TFPK^\alpha L^{1-\alpha}$. We find: [20]

$$\begin{aligned} K^{1-\alpha} =& TFPL^{1-\alpha}\left(\frac{P(1-\alpha)}{w}\right)^{\frac{\eta-\alpha\eta}{1-\eta}} \alpha^{\frac{1+\alpha\eta-\eta}{1-\eta}} \left(\frac{a_L^\eta}{\mu_L r_L^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}} \\ &\left[Prob(a_{it} = a_L) + \sum_{\tau=1}^{\infty} Prob(a_{it} = a_H \& s = t-\tau)G_\tau\right] \end{aligned} \tag{29}$$

2.5 Labor Market clearing

Since there are two markets, by Walras' Law, general equilibrium is realized when the labor market clears. Labor demand, given in equation (18), from all firms has to equal labor supply L:

$$L = \sum_{i=1}^{M}\left(\frac{(1-\alpha)}{\mu_{it}}\frac{P}{w}\left(\frac{Q}{M}\right)^{1-\eta} a_{it}^\eta k_{it}^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}$$

In appendix A.2, this equation is derived further. Then, notice that labor market clearing is realized for the following $\frac{P}{w}$:

$$\frac{P}{w} = \left(\frac{L}{K}\right)^\alpha \frac{\Omega^{\eta-\alpha\eta-1}}{(1-\alpha)\left(\frac{TFP}{M}\right)^{1-\eta}} \tag{30}$$

where

$$\Omega \equiv \left[\sum_{i=1}^{M}\left(\frac{a_{it}^\eta}{\mu_{it}}\left(\frac{\left(\frac{a_{it}^\eta}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M}\left(\frac{a_{it}^\eta}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}\right)^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}\right] \tag{31}$$

Like TFP, Ω is a function of the joint distribution of productivities, markups and capital. In a context with monopolistic competition, i.e. without variable markups, this condition would not exist.

[20] Specifically:

$$K = Q\left(\frac{P(1-\alpha)}{w}\right)^{\frac{\eta-\alpha\eta}{1-\eta}} \alpha^{\frac{1+\alpha\eta-\eta}{1-\eta}} \left(\frac{a_L^\eta}{\mu_L r_L^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}} \left[Prob(a_{it} = a_L) + \sum_{\tau=1}^{\infty} Prob(a_{it} = a_H \& s = t-\tau)G_\tau\right]$$

In short, the above implies that the labor market clearing equation can be written as:

$$\frac{P}{w} = F_L(M, L, K, TFP, \Omega) \tag{32}$$

2.5.1 Summary of steady-state equilibrium

The nature of the steady-state equilibrium will be determined by the following elements:

- $H(a, k)$, the joint distribution of a_{it}, k_{it}, characterized in Lemma 1
- The distribution of markups, characterized in equations (23), (24)
- Aggregate TFP, characterized in (28)
- Aggregate capital, characterized in (29)
- The factor-price ratio, determined in the labor-market-equilibrium condition in (32)
- Ω, characterized in (31).

In the comparative-statics exercise that now follows, I describe how the steady-state variables change with M. A crucial role there will be played by the comparative statics on G_τ, which is a crucial determinant of the distribution of capital.

2.6 Comparative statics on competition

In the theoretical appendix sections, I demonstrate the following proposition on the comparative statics for M:

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

- *Markup levels fall with M:*

$$\mu'_L < \mu_L\,;\, \mu'_H < \mu_H\,;\, \mu'_\tau < \mu_\tau$$

- *Markup dispersion falls with M:*

$$\frac{\mu'_H}{\mu'_L} < \frac{\mu_H}{\mu_L}\,;\, \frac{\mu'_\tau}{\mu'_L} \leq \frac{\mu_\tau}{\mu_L}$$

- *Capital wedges worsen with M:*

$$\omega'_\tau \geq \omega_\tau$$

and $(\Phi_\tau > 0) \implies (\omega'_\tau > \omega_\tau)$

The proposition demonstrates the dual role of competition in an environment with both variable markups and financial constraints. On the one hand, misallocation due to markup distortions improves, since both markup levels and markup dispersion fall with M. On the other hand, misallocation due to capital wedges worsens due to competition. Since the latter effect is absent in a setting without financial constraints while the former is not, the welfare gains from competition tend to be lower in a setting with financial constraints compared to a setting without financial constraints.

In what follows, I will refer to the combined decrease in markup levels and markup dispersion due to competition, as a reduction in *markup misallocation*.[21] When capital wedges worsen for all financially constrained firms, I will summarize this as an increase in *capital misallocation*.

3 Data on Indian manufacturing plants

To test the predictions of the model, the empirical analysis employs establishment-level panel data from the Indian Annual Survey of Industries (ASI), for the period 1990-2011. The ASI sampling scheme consists of two components.[22] One component is a census of all manufacturing establishments with more than 100 employees, while a second component samples, with a certain probability, each formally registered establishment (or plant) with less than 100 employees. All establishments with more than 20 workers (10 workers if the establishment uses electricity) are required to be formally registered.[23]

In the empirical exercise, I will be exploiting variation across 3-digit sectors[24] and geographical units in India. The geographical units in the data are either states or union territories. For convenience, I will be referring to both geographical units as "states."[25]

The main plant-level variables used in the analysis are capital K_{irst}, labor L_{irst}, materials M_{irst} and revenue S_{irst}, for plant i, state r, sector s and year t. Here, a year is defined as the financial year, and K_{irst} is the book value of assets at the start of the financial year. The logarithm of a variable will be denoted in lower case.

4 Motivating stylized facts

Before examining the causal predictions of the model, I first present two motivating stylized facts. A first stylized fact describes the empirical association of productivity volatility and measured capital misallocation, which corroborates the central role for productivity volatility in the model. Second, I document a negative macro-level empirical relationship between the level of markups and measured capital misallocation. This macro-level stylized fact is in line with the main predictions of the model, and serves to set the stage for the plant-level analysis

[21]Note that reduced markup levels improve intertemporal allocative efficiency of the composite output good, by inducing higher aggregate saving and investment.

[22]The particulars provided here hold for the majority of the sample years. Bollard et al. (2013) provide a more detailed description of the ASI data, including certain modifications to the sampling scheme.

[23]For the years 1998-2011, establishment identifiers are provided by the Indian Statistical Office. For the pre-1998 years, I use the panel-identifiers employed by Allcott et al. (2014), which were generously made available by Hunt Allcott.

[24]For all the empirics related to India's 1997 dereservation reform, sector definitions are based on the 2004 National Industrial Classification (NIC). For all other empirical exercises, the 1987 classification is used.

[25]To make the definitions of states consistent over time, I employ the concordance provided by the Indian Statistical Office. This results in a number of 35 states in the panel data.

of the causal role of competition in Proposition 1.

4.1 Productivity volatility drives capital misallocation

A central mechanism in my model is how capital misallocation arises from the interaction of financial constraints with firms' idiosyncratic productivity shocks. I now check if the theoretical role of productivity volatility as a driver of capital misallocation is empirically relevant for the Indian manufacturing sector. Crucially, the goal will not be to establish a causal link between productivity volatility and capital misallocation. Instead, I examine if there indeed exists a correlation between productivity volatility and capital misallocation, which is a necessary condition for the empirical relevance of my model.

To document the empirical link between productivity volatility and capital misallocation, I replicate the analysis of Asker et al. (2014) - henceforth ACWDL - for the ASI data of the Indian manufacturing sector. As in ACWDL, the analysis centers around dispersion of marginal revenue product of capital (MRPK). The measure for MRPK is based on the assumption of a sector-level Cobb-Douglas production function, which implies that the marginal revenue product of capital takes the following form:

$$MRPK_{irst} = \ln(\beta_s^K) + s_{irst} - k_{irst} \tag{33}$$

MRPK dispersion is then measured at the sector-year level as $Std_{st}(MRPK_{irst})$, and the empirical measure for productivity volatility is $Std_{st}(a_{it} - a_{it-1})$, with a_{it} as the measure of plant-level revenue productivity.[26] [27]

The data exhibits a strong upward-sloping relationship between productivity volatility and MRPK dispersion (Figure 1). Moreover, as in ACWDL, I also implement several plant-level robustness tests for the relationship between MRPK dispersion and productivity volatility (see Appendix F). This combination of sector-level evidence with plant-level evidence on the correlation between productivity volatility and capital misallocation, substantially reduces the likelihood that this correlation is only driven by measurement error. Therefore, this evidence strongly corroborates the empirical relevance of productivity volatility as a driver of capital misallocation.

[26] Revenue productivity is again measured as in ACWDL, who impose that revenue takes a Cobb-Douglas form. Together with the assumption of cost-minimization these structural assumptions imply that productivity a_{it} can be measured as:

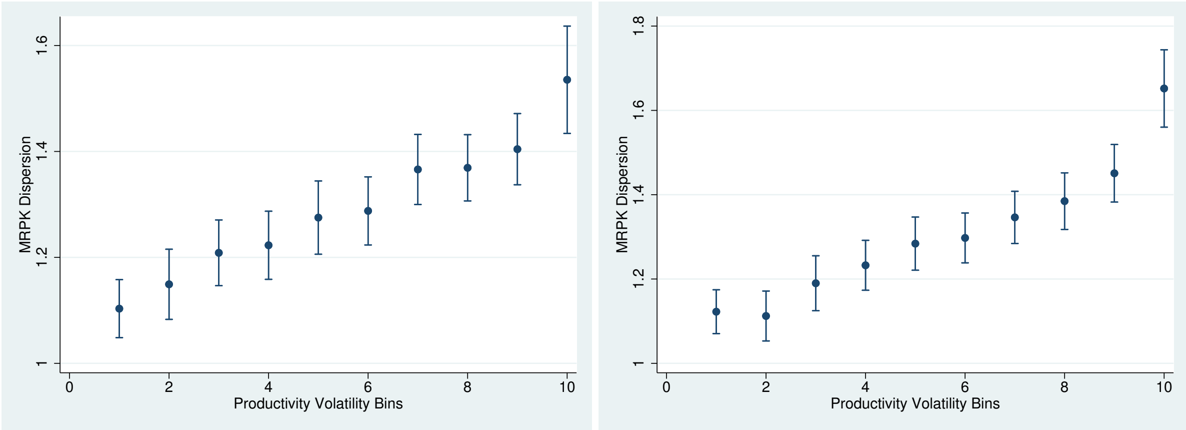
$$a_{it} = s_{irst} - \beta_s^K k_{it} - \beta_s^L l_{it} - \beta_s^M m_{it}$$

where output elasticities are measured as $\beta_s^L = Median_s\left[\frac{wage\ bill_{irst}}{S_{irst}}\right]$, $\beta_s^M = Median_s\left[\frac{M_{irst}}{S_{irst}}\right]$, $\beta_s^K = 1 - \beta_s^L - \beta_s^M$. To avoid sensitivity to outliers, the median is calculated at the 2-digit sector level.

[27] Throughout, I also employ measures based on value-added as a leading robustness check. Employing different productivity measures based on either gross revenue or value added serves as a primary robustness check. Since the measured elasticities for labor and capital are meaningfully different in the two measures, any sensitivity of the findings to the particular choice of output elasticities is substantially mitigated. Empirically, value added is measured as $VA_{irst} = S_{irst} - M_{irst}$, $MRPK_{irst}^{VA} = \ln(\beta_s^{K,VA}) + va_{irst} - k_{irst}$ and $a_{irst}^{VA} = va_{irst} - \beta_s^{K,VA} k_{irst} - \beta_s^{L,VA} l_{irst}$, with $\beta_s^{L,VA} = Median_s\left[\frac{wage\ bill_{irst}}{VA_{irst}}.\right]$; $\beta_s^{K,VA} = 1 - \beta_s^{L,VA}$

Figure 1: Correlation between productivity volatility and MRPK dispersion

(a) Gross revenue based (b) Value-added based



MRPK dispersion is measured as $Std_{st}(MRPK_{irst})$ and productivity volatility as $Std_{st}(a_{it} - a_{it-1})$. The figure shows point estimates and 95% confidence intervals for γ_D from the regression $Std_{st}(MRPK_{irst}) = \sum_{D=1}^{10} \gamma_D 1(Decile\ D)_{st} + \varepsilon_{st}$. Here, $1(Decile\ D)_{st}$ indicates if a sector-year observation belongs to decile D of the distribution of $Std_{st}(a_{it} - a_{it-1})$. $MRPK_{irst}$ is measured as in equation (33), and measurement of productivity a_{it} is described in footnote 26. The value-added based measures for panel (b) are described in footnote 27.

Note that this analysis does not claim that financial constraints are the sole mechanism that leads to the strong correlation between measured productivity volatility and measured capital misallocation in the Indian manufacturing sector. For instance, ACWDL propose the alternative mechanism of capital adjustment-costs as an explanation for this pattern. Another theoretically possible mechanism is that productivity volatility induces volatility in markups, which would also be captured in the ACWDL measure of capital misallocation. Importantly, both of the above mechanisms assign a crucial role to productivity volatility in generating misallocation, which supports the empirical relevance of bringing productivity volatility into the theory. Moreover, standard models do not predict adjustment costs or markup volatility to increase as competition intensifies. Since the following sections will show that capital wedges increase with competition, this is further evidence in favor of the interaction of productivity volatility with financial constraints as a central empirical mechanism for generating misallocation.

4.2 Correlation between Competition and Misallocation

Proposition 1 states that, when the number of firms increases, the level of markups decreases for all types of firms and capital wedges increase for financially constrained firms. Hence, a direct implication of the proposition is that, ceteris paribus, lower markups are associated with increased capital misallocation. I now show that this correlation between markup levels and capital misallocation is indeed present in the data. To be clear, this evidence should not be interpreted as causal. Instead, the examination of the correlation between markups and competition is a first check of the consistency of the data with the prediction of the model on capital misallocation.

Proposition 1 implies that all first moments of the distribution of markups fall as competition increases. In this analysis I therefore focus on the median markup $Median_{rst}[\ln \mu_{irst}]$, which is a robust first moment of the markup distribution. Here, μ_{irst} is the plant-level markup, measured as in De Loecker and Warzynski (2012). This measurement is based on the assumptions that plants have Cobb-Douglas production functions and minimize costs, and that labor is a variable input. As explained in appendix E, these structural assumptions lead to the following expression for μ_{irst}:

$$\mu_{irst} = \beta_s^L \frac{VA_{irst}}{w_{irst} L_{irst}} \tag{34}$$

where $w_{irst}L_{irst}$ is the wage bill and β_s^L is the sectoral output-elasticity of labor. Intuitively, when plants spend a higher share of value added on labor, conditional on the output elasticity for labor, these firms are setting a lower markup.

The measure for capital misallocation is still the ACWDL measure for MRPK dispersion. In a regression analysis, I then employ the following specification:

$$Std_{rst}(MRPK_{irst}) = \gamma_s + \gamma_t + \gamma_r + \zeta Median_{rst-1}[\ln \mu_{irst-1}] + \epsilon_{rst} \tag{35}$$

where $\gamma_s, \gamma_t, \gamma_r$ are sector, year and state fixed effects respectively. In alternative specifications, I also run this regression without γ_t or γ_r. However, I always include sector fixed-effects γ_s to eliminate variation arising from β_s^L, the sectoral output elasticity for labor.

Results Table 1 provides suggestive evidence for the prediction that MRPK dispersion increases with competition. First we notice that $Std_{rst}(MRPK_{irst})$ is consistently negatively related to the median markup in a state-sector-year observation. This holds for both gross-revenue and value-added based measures of $MRPK$, and it holds regardless of the specific set of fixed effects.[28]

[28] One might be worried about a mechanical correlation between the level of $Median_{rst-1}[\mu_{irst-1}]$ and the level of $Std_{rst}(MRPK_{irst})$. Note, however, that this would imply a positive correlation, while the regressions in Table 1 demonstrate a persistently negative correlation.

Table 1: MRPK Dispersion and Competition

	$Std_{rst}(MRPK_{irst}(Gross\ Revenue))$				$Std_{rst}(MRPK_{irst}(Value\ Added))$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Median_{rst-1}[\ln \mu_{irst-1}]$	-0.0547**	-0.0494**	-0.0371**	-0.0353**	-0.0501**	-0.0447**	-0.0376**	-0.0353**
	(0.0102)	(0.0101)	(0.0102)	(0.0101)	(0.00977)	(0.00972)	(0.0101)	(0.00997)
Constant	1.306**	1.236**	1.257**	1.228**	1.321**	1.438**	1.336**	1.477**
	(0.00487)	(0.0125)	(0.0329)	(0.0347)	(0.00465)	(0.0291)	(0.0362)	(0.0457)
Year FE	No	Yes	No	Yes	No	Yes	No	Yes
State FE	No	No	Yes	Yes	No	No	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	19951	19951	19951	19951	19570	19570	19570	19570

Standard errors, clustered at the state-sector level, in parentheses (* $p < 0.05$, ** $p < 0.01$).
Indices are i for plant, r for state, s for sector and t for year. Hence, observations are at the state-sector-year level.
Specifications 1-4 measure MRPK based on gross revenue, and specifications 5-8 based on value added.

The above evidence suggests that increased competition is associated with macro-level capital misallocation. Since this association lacks causal identification, the next empirical section will aim to establish a negative causal relationship between competition and plant-level capital wedges at the micro level.

5 Natural experiment: a competition policy reform

Proposition 1 predicts a dual effect of competition: it reduces markup misallocation and increases capital misallocation. In this section, I will use a natural experiment - India's dereservation reform - to test the theory's predictions on the impact of competition at the plant level.

5.1 Background on the dereservation reform

To understand how dereservation creates natural variation in the degree of competition, I start by describing the specifics of this 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 investment ceiling (10 million Rupees at historical cost in 1999) were allowed to produce certain product categories.[29] In 1996, before the start of dereservation, around 1000 product categories were reserved for SSI.

Starting in 1997, the Indian government starts with gradually removing the reservation policy, and the process of dereservation peaks between 2002 and 2008.[30]. Interestingly, Tewari and Wilde (2016) 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 the same demand and supply characteristics, this limits the scope for a structural explanation of the timing of dereservation. This

[29] 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.

[30] A detailed description of the history of reservation policies and of the implementation of dereservation is provided by García-Santana and Pijoan-Mas (2014); Martin et al. (2014) and Tewari and Wilde (2016).

is important, since the exogeneity of the timing of dereservation is crucial for my identification strategy, and I discuss how to test this exogeneity in greater detail below.

5.1.1 Dereservation as a pro-competitive reform

The practical implications of the dereservation reform have strong structural similarities with the model's definition of a pro-competitive shock. To understand why, I first explain what, from an ex-ante theoretical point of view, the main possible effects of deservation are. The plant-level effects of dereservation will vary depending on whether a plant was producing reserved product categories prior to dereservation or not. Dereservation has two distinct structural effects on incumbent plants, which are defined as plants whose main product was reserved prior to dereservation. First, the direct effect of the removal of the investment ceiling is that incumbents are allowed to grow their capital stock. Second, there is the pro-competitive shock from dereservation on incumbents. The removal of the reservation policy implies that any plant is now allowed to produce the previously reserved product. As a result, there is substantial scope for entry into the production of dereserved products by non-incumbent plants.

If dereservation indeed leads to a higher number of firms in the market, this dimension of the reform corresponds directly to the comparative statics in my model. Therefore, the staggered implementation of the dereservation reform provides a natural experiment to test the model's predicted effects of an increase in competition. Does higher competition reduce markup levels and increase capital wedges, as predicted by Proposition 1?

5.1.2 Data on the dereservation reform

Data on the dereservation reform has been generously provided by Ishani Tewari, and a complete description of this data and its construction is available in Tewari and Wilde (2016). I will define a plant as being dereserved in year t if that plant's main product has been deserved during that financial year.

5.1.3 Existing evidence on the dereservation reform

Before testing Proposition 1 with the dereservation reform, it is useful to gain a better understanding of the dereservation reform by summarizing the existing evidence on this reform.

Timing of dereservation Some of the regression specifications below will rely on a difference-in-difference type identification. For such an analysis, it is crucial that the timing of dereservation is exogenous, and is therefore orthogonal to pre-dereservation trends. Martin et al. (2014) examine pre-dereservation trends by year of dereservation, and find no evidence for any difference in trends. This precise null-result holds for labor growth, as well as for other outcome measures.

Pro-competitive impact In addition to testing for the exogenous timing of the dereservation reform, Martin et al. (2014) also provide evidence for the pro-competitive impact of the reform. More specifically, they find that dereservation "led to the entry and expansion of output, employment and investment among new entrants to the previously reserved product space." At the same time, the market shares of incumbent plants fall. These findings have a strong similarity with an increase in competition in my model; the number of firms in a market increases such that market shares fall.

5.2 Impact of dereservation on markups

Event study design I now describe the empirical strategy to test if the dereservation reform leads to lower markup levels. In this test, I measure μ_{irt}, the markup for plant i - located in state r - in year t, by using the method from De Loecker and Warzynski (2012), who show that:[31]

$$\mu_{irt} = \beta^L_{ir} \frac{VA_{irt}}{w_{irt} L_{irt}}$$

where $\frac{w_{irt} L_{irt}}{VA_{irt}}$ is labor's share of value added, and β^L_{ir} is the output elasticity of labor.[32] Using this markup measure, I run the following event-study on dereservation, where I define the time at which the main product of plant i is dereserved as e_{irt}.

$$\ln \mu_{irt} = \alpha_{ir} + \gamma_{rt} + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{irt} + \tau] + \varepsilon_{irt} \tag{36}$$

Here, α_{ir} is a plant fixed effect and γ_{rt} is a state-year fixed effect. I also bin up the end-points and normalize $\beta_{-1} = 0$. For the purpose of this event study, I restrict the sample to a balanced sample of incumbent plants, which includes all plants whose main product category was reserved prior to dereservation.[33] In this specification, as in all the following, standard errors are clustered at the plant level.

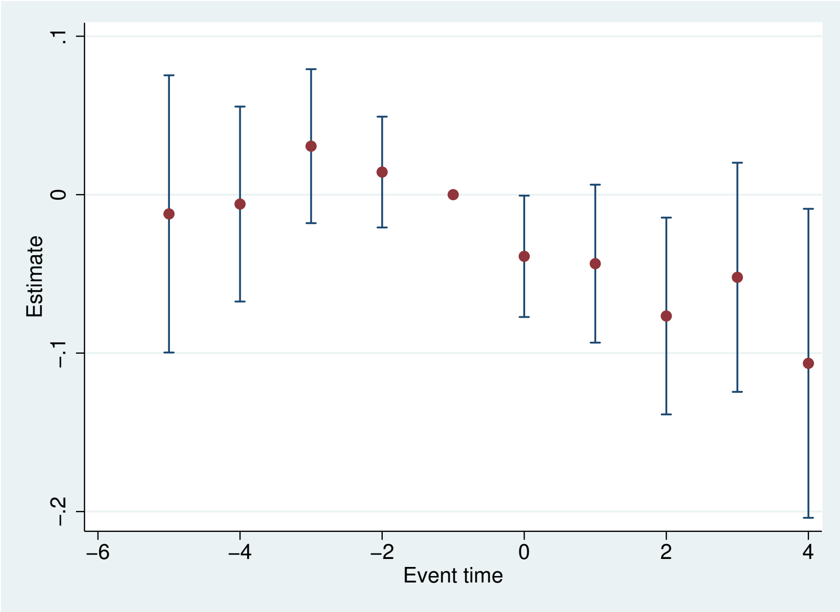
Reduced markup levels Figure 2 demonstrates that markups fall due to the dereservation reform. As soon as dereservation is implemented, there is an immediate and significant decline in markups. In subsequent periods, this fall in markups tends to grow stronger in terms of economic magnitude. Hence, the impact of dereservation is in line with the theoretical prediction that markup levels fall as the number of firms increases. Moreover, the economic magnitude of the impact of the reform is substantial, since four years after dereservation, the

[31] As explained in Appendix E, this method relies on the assumptions of Cobb-Douglas production functions, cost minimization and labor as a variable input.

[32] This expression is similar to equation (34), except that I now allow for plant-level variation in the output elasticity. In the regression analysis, a plant-level fixed effect will absorb this output elasticity.

[33] Another option would be to also include non-incumbent plants, but these plants are not necessarily a valid control group.

Figure 2: Dereservation Event-study on Markups



The figure displays the coefficients and 95% confidence intervals for the β_τ coefficients from the following event-study regression: $\ln \mu_{irt} = \alpha_i + \gamma_{rt} + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{irt} + \tau] + \varepsilon_{irt}$. Here, α_i is a plant fixed-effect and γ_{rt} is a state-year fixed-effect. I define the time at which a first product of plant i is dereserved as e_{irt}. I impose the normalization that $\beta_{-1} = 0$, and cluster standard errors at the plant-level.

average markup of an incumbent plant has declined by 0.11 log points.[34]

Markup dispersion In addition to leading to a decrease in 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 a plants' markup before dereservation is above or below the median markup. I then find that plants who have higher markups in the periods before dereservation, exhibit a stronger decline in their markup after dereservation, as predicted by the theory (see Appendix G for details).[35]

[34]Note also that there is no evidence of a downward trend in markups prior to dereservation, since the estimates of β_τ for $\tau < -1$ are all close to zero, both statistically and economically.

[35]For the subset of plants with below-median initial markup, I cannot reject that dereservation has no effect on markup levels, whereas the effect for plants with above-median initial markups is large and strongly significant. For the latter group, there appears to be a significant change in the markup at $\tau = -1$. This is potentially due to anticipation effects. Since the final phase of the government's decision process about the dereservation of a product category involved consultation with potential stakeholders, it may be possible for both incumbent or entrant plants to anticipate dereservation in the immediate pre-period. In any case, the decline in markups for this group accelerates after dereservation, and therefore the strong negative impact of dereservation on markup levels for this group of plants is robust.

5.3 Dereservation and capital misallocation

5.3.1 Empirical tests for capital convergence

After testing the impact of dereservation on markup levels, I now examine the impact of the dereservation reform on capital wedges. In the model, firms optimally choose to grow their capital stock in response to positive productivity shocks until they reach k_H^*, the optimal level of capital for high productivity firms (implied by the first-order condition from equation (14)). The empirical challenge here is that k_H^*, and therefore a firm's capital wedge, is unobserved.

My empirical strategy to test for competition's effect on capital wedges is inspired by Asker et al. (2014) (or ACWDL), who focus on convergence in terms of marginal revenue product of capital (MRPK). From this perspective, the inability for a financially constrained firm to reach its optimal capital level at time t, implies that for this firm i, $MRPK_{it}^* < MRPK_{it}$. Here $MRPK_{it}$ is firm i's actual MRPK in period t, and $MRPK_{it}^*$ is its optimal MRPK from the unconstrained solution. Since $MRPK_{it}$ is a strictly monotone function of k_{it}, and capital convergence in the model slows down with M, convergence in terms of MRPK also slows down with M. Given this restatement of the model's prediction on capital wedges, the empirical strategy now requires a measure for $MRPK_{it}$ and especially for the more challenging $MRPK_{it}^*$. While $MRPK_{it}^*$ will remain unobservable, the theory predicts it to be substantially more stable over time than k_{it}^*, which is why the shift in focus from capital to MRPK is useful.

My measure for MRPK is very similar to the one in equation (33), except that now I allow for plant-level variation in the output elasticity of capital β_i^K.

$$MRPK_{it} = \ln(\beta_i^K) + s_{it} - k_{it} \tag{37}$$

In the regressions, I will use plant-level fixed effects to absorb β_i^K, so within-plant variation in MRPK will be solely driven by the log difference between revenue and capital. Next, I will describe how to analyze the impact of dereservation on convergence to $MRPK_{it}^*$, as well as how to proxy for $MRPK_{it}^*$.

5.3.2 Econometrics

To examine if MRPK convergence slows down due to dereservation, I use the following autoregressive framework:

$$\begin{aligned} MRPK_{it} = {} & \alpha_i + \gamma_t + \beta_1 Deres_{it-1} + \rho_0 MRPK_{it-1} \\ & + \rho_1 MRPK_{it-1} * Deres_{it-1} + \beta_2 \ln age_{it} + \varepsilon_{irt} \end{aligned} \tag{38}$$

where α_i and γ_t are plant and year fixed-effects respectively. The main coefficient of interest in this specification is ρ_1. This coefficient estimates how the speed of convergence to $MRPK_{it}^*$ changes as a function of dereservation. To better understand the estimation strategy, as well as the measurement of $MRPK_{it}^*$, 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, we will find that the average plant experiences a delayed adjustment to $MRPK^*_{it}$. Crucially then, if $\rho_0 > 0$, then $\rho_1 > 0$ will indicate that the speed of MRPK convergence slows down with dereservation. In equation (38), I proxy for $MRPK^*_{it}$ with $\alpha_i + \gamma_t + \beta_1 Deres_{it-1} + \beta_2 \ln age_{it}$, but the estimation results for ρ_1 will be robust to the exact choice for the proxy.

Equation (38) will be estimated on a sample with only incumbents, i.e. plants who were producing reserved products prior to dereservation. I also estimate the impact of dereservation on the full sample of plants, for two reasons. First, the sample of incumbent is relatively small, which limits the statistical power of my autoregressive framework. Second, the full sample of plants allows me to control for local economic shocks at the state-sector-year level, which is impossible when only examining incumbents, due to collinearity issues.[36]

5.3.3 Results on capital convergence

Table 2 presents the impact of dereservation on MRPK convergence, and these results are in line the theoretical predictions of the model. First, the theory requires that $0 < \rho_0 < 1$, as there is convergence to $MRPK^*_{it}$, but this convergence is not immediate due to financial constraints. In all specifications, the 95% confidence interval is always well within the $[0, 1]$ interval. The point estimates for ρ_0 are generally below 0.5, which implies that convergence to $MRPK^*_{it}$ is relatively fast. Hence, the proxy for $MRPK^*_{it}$ appears to be empirically valid.[37]

Importantly, all coefficients on the interaction of dereservation with $MRPK_{irst-1}$ are positive, as predicted by the theory, and in 3 of the 4 specifications, the coefficients are strongly statistically significant. The estimated magnitude of the effect of dereservation is modest but economically meaningful. For specifications 1 and 2 specifically, dereservation increases the

[36]Collinearity issues arise from small numbers of incumbent plants in many state-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 "stayer" - includes all remaining plants. For this full sample of plants, I employ the following estimation specification:

$$\begin{aligned} MRPK_{irst} = {} & \alpha_{irs} + \gamma_{rst} + \beta_1 Deres_{irst-1} + \beta_2 Deres_{irst-1} * entrant_{irs} \\ & + \rho_0 MRPK_{irst-1} + \rho_1 MRPK_{irst-1} * incumb_{irs} + \rho_2 MRPK_{irst-1} * entrant_{irs} \\ & + \rho_3 MRPK_{irst-1} * Deres_{irst-1} + \rho_4 MRPK_{irst-1} * Deres_{irst-1} * entrant_{irs} \\ & + \beta_3 X_{irst} + \varepsilon_{irst} \end{aligned} \tag{39}$$

Here, α_{irs} is a plant fixed-effect, $entrant_{irs}$ and $incumb_{irs}$ are indicators for plant i being entrants or incumbents, and γ_{rst} is a state-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 levels post dereservation for incumbents and entrants. The middle row estimates convergence speeds prior to dereservation, allowing for different speeds of convergence for stayers, incumbents and entrants. 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.

[37]I examined different variations of the proxy for MRPK. Empirically, the strongest factor in increasing convergence speed (lowering ρ_0 closer to 0), is the plant-level fixed effect. From a theoretical point of view, plant-level variation in $MRPK^*_{it}$ could be driven by variation in interest rates across plants, for instance due to different risk profiles.

Table 2: Dereservation and MRPK convergence

	$MRPK_{it}$ - Value Added (VA)		$MRPK_{it}$ - Gross Revenue (GR)	
	(1)	(2)	(3)	(4)
$Deres_{it-1}$	0.190***	0.196***	0.127***	0.185***
	(0.0545)	(0.0544)	(0.0363)	(0.0369)
$MRPK_{it-1}(VA)$	0.312***	0.221***		
	(0.00931)	(0.00685)		
$MRPK_{it-1}(VA) * Deres_{it-1}$	0.0634***	0.0418***		
	(0.0132)	(0.0129)		
$MRPK_{it-1}(GR)$			0.425***	0.348***
			(0.0106)	(0.00783)
$MRPK_{it-1}(GR) * Deres_{it-1}$			0.0638***	0.0578***
			(0.0122)	(0.0119)
Plant Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	Yes	No
State-sector-year Fixed Effects	No	Yes	No	Yes
Observations	62048	173931	69534	204112

P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In specifications 1 and 2, MRPK is measured based on value added, and based on gross revenue in specifications 3 and 4. Specifications 1 and 3 estimate equation (38) on a sample restricted to all plants that were incumbent more than 2 years before their main product was dereserved. Specifications 2 and 4 estimate equation (39) on the full sample. All specifications control for the logarithm of a plant's age. Standard errors are clustered at the plant level.

half-life of the autoregressive process by respectively 18.9% and 13%.[38] [39]

Due to data constraints, I can only observe the impact of dereservation over a limited time frame. This imposes a limitation for testing the predictions of my model. After all, the comparative statics in my model are across steady states, whereas the empirical analysis of dereservation will also capture transitory dynamics. To address this concern, in the next section I extend the analysis to the full sample, where it may be more valid to assume that markets are in steady state.

6 Competition and MRPK convergence in the full panel

6.1 Empirical strategy in the full panel

To provide further evidence for the predictions of the model, I will now test the prediction for capital convergence on the full sample. My empirical strategy will use a measure of competition that is arguably exogenous from the point of view of the individual plant, which allows

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

[39]One concern for my estimation strategy is the bias described by Nickell (1981), which leads to a downward bias on ρ_0. This helps explain why the estimated values for ρ_0 are low. Note that if ρ_1 has a downward bias, then this works against finding evidence for dereservation slowing down MRPK convergence.

me to analyze MRPK convergence at the plant level. Unfortunately, this empirical strategy will not allow me to test the impact of competition on markup misallocation. Since the prediction on capital convergence is the most novel prediction of my model, and the predictions on markup misallocation have been examined in previous research (see e.g. Peters (2013), Schaumans and Verboven (2015)), it makes sense to focus on examining capital convergence in closer detail.

Measuring competition I will use the median markup at the state-sector-year level, namely $Median_{rst}[\ln \mu_{irst}]$ (measured as in equation (34)), as an inverse measure of competition. Since this competition measure is arguably exogenous from the plant's point of view, this allows me to examine the causal link between competition and MRPK convergence at the plant-level.[40]

Estimation strategy To implement the empirical test on MRPK convergence, I update the autoregressive framework from equation (38) in the following way.

$$\begin{aligned} MRPK_{irst} &= \alpha_{irs} + \gamma_{rst} + \beta \ln age_{irst} + \rho_0 MRPK_{irst-1} \\ &\quad + \rho_1 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] + \varepsilon_{irst} \end{aligned} \tag{40}$$

As before, the main coefficient of interest is ρ_1. This coefficient estimates how the speed of convergence changes as a function of $Median_{rst-1}[\ln \mu_{irst-1}]$.[41] The econometrics section for the dereservation reform contains a detailed explanation of this type of autoregressive specification (see Section 5.3.2). Remember 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, the typical plant will experience a 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 $Median_{rst-1}[\ln \mu_{irst-1}]$.[42]

6.2 Heterogeneity along financial dependence

So far, my tests on capital convergence have all implicitly assumed that the average plant in the sample is financially constrained. However, this implicit assumption masks the empirical heterogeneity in the degree to which plants are financially constrained. To address this concern,

[40] From the point of view of the model, an alternative measure would be M_{rst}/L_{rt}, i.e. the number of firms per capita. Note however that in the model there is a monotone relationship between, first M_{rst}/L_{rt}, second the first moments of the marketshare distribution, and third the first moments of the markup distribution. The advantages of the median markup are therefore that it has a direct link to M_{rst}/L_{rt}, i.e. the exogenous degree of competition in the model, and that it incorporates factors left out of the model, such as variation in market size coming from sectoral expenditure shares or income per capita.

[41] Since $Median_{rst-1}[\ln \mu_{irst-1}]$ enters in an interaction term, in the regressions it is demeaned, and normalized to standard deviation units. Demeaning happens within sectors, to avoid results being driven by variation in the measurement of the output elasticity β^L_s.

[42] A sufficient condition for the speed of MRPK convergence to increase with $Median_{rst-1}[\ln \mu_{irst-1}]$ is that $|\rho_0 + \rho_1 * Median_{rst-1}[\ln \mu_{irst-1}]| < \rho_0$.

I now explore the implications of heterogeneity along financial dependence for MRPK convergence. Following Rajan and Zingales (1998), the idea is that for 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 MRPK convergence.

Specifically, sectoral financial dependence is measured as

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

based on data for US sectors over the entire 1980's.[43] 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 US. The central idea in Rajan and Zingales (1998) is then that in a setting such as India, with less developed financial markets, financial constraints become especially binding in sectors with high levels of $Fin\ Dep_s$.

Estimation strategy To examine the role of financial dependence in the setting of MRPK convergence, I augment the earlier specification to allow for heterogeneous effects along financial dependence:

$$\begin{aligned} MRPK_{irst} = \alpha_{irs} + \gamma_{rst} + \beta \ln age_{irst} + \rho_0 MRPK_{irst-1} \\ + \rho_1 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] + \rho_2 MRPK_{irst-1} * Fin\ Dep_s \\ + \rho_3 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] * Fin\ Dep_s + \varepsilon_{irst} \end{aligned} \quad (41)$$

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.

6.3 Estimation results for the full sample

The results for MRPK convergence in the full sample (Table 3) confirm the results from the analysis of dereservation. First, across all specifications, MRPK converges strongly to the empirical proxy for $MRPK^*_{irst}$, but this convergence is not immediate. Formally, for all conventional levels of statistical significance, $0 < \rho_0 < 1$.

Second, the speed of convergence always increases with $Median_{rst}[\ln(\mu_{irst-1})]$, the inverse measure for competition, for baseline specification (40). Specifically, the coefficient on ρ_1 is always negative and strongly statistically significant with p-values below 0.01 (see columns 1,2,5,6). 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, just as in the case of dereservation. As an example, in specification 2, an increase in the

[43] I use the original Rajan and Zingales (1998) measures of financial dependence for ISIC Rev.2 sector definitions. These 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.

median markup by two standard deviations, decreases the half-life of MRPK convergence by 5.7%.

The results for heterogeneity along financial dependence are also in line with expectations (see columns 3,4,7,8). The coefficient ρ_3, estimated on the triple interaction term, is significantly negative in three of the four specifications. The one exception is the estimate in column 4, which is statistically insignificant.[44] The estimation result that $\rho_3 < 0$ implies that the magnitude of the influence of the median markup increases with financial dependence. Consider for instance the industry producing electric machinery, which has a relatively high level of financial dependence (measured at 77% by Rajan and Zingales (1998).) For this sector, an increase in the median markup by two standard deviations, decreases the half-life of MRPK convergence by 7.2%.

[44]Note that the coefficient on $MRPK_{irst-1} * FinDep_s$ is always positive, which is consistent with MRPK convergence being slower in more financially dependent sectors.

Table 3: Competition and Speed of MRPK Convergence

	$MRPK_{irst}$ (Value added (VA))				$MRPK_{irst}$ (Gross Revenue (GR))			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$MRPK_{irst-1}(VA)$	0.264***	0.225***	0.244***	0.195***				
	(0.00609)	(0.00605)	(0.00690)	(0.00737)				
$MRPK_{irst-1}$ (VA) * $Median_{rst-1}[\ln \mu_{irst-1}]$	-0.00824***	-0.00941***	0.000227	-0.00657*				
	(0.00175)	(0.00260)	(0.00251)	(0.00379)				
$MRPK_{irst-1}$ (VA) * $Fin\ Dep_s$			0.0322***	0.0329***				
			(0.00830)	(0.00892)				
$MRPK_{irst-1}$ (VA) * $Median_{rst}[\ln(\mu_{irst})] * Fin\ Dep_s$			-0.0224***	0.00205				
			(0.00547)	(0.00874)				
$MRPK_{irst-1}(GR)$					0.397***	0.382***	0.379***	0.359***
					(0.00673)	(0.00645)	(0.00777)	(0.00793)
$MRPK_{irst-1}$ (GR) * $Median_{rst-1}[\ln \mu_{irst-1}]$					-0.00843***	-0.00582**	-0.00532*	0.000512
					(0.00213)	(0.00274)	(0.00271)	(0.00386)
$MRPK_{irst-1}$ (GR) * $Fin\ Dep_s$							0.0169*	0.00879
							(0.00963)	(0.0105)
$MRPK_{irst-1}$ (GR)* $Median_{rst}[\ln(\mu_{irst}] * Fin\ Dep_s$							-0.0139**	-0.0227**
							(0.00649)	(0.00954)
Plant Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	Yes	No	Yes	No	Yes	No
State-sector-year Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	177475	145363	142239	113856	216250	180370	172597	140951

P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In specifications 1 - 4, MRPK is measured based on value added, and based on gross revenue in specifications 5-8. All specifications control for the logarithm of a plant's age. Standard errors are clustered at the plant level. The variable $Median_{rst-1}[\ln(\mu_{irst-1}]$ is demeaned within sectors and measured in standard deviation units.

7 Additional evidence from undercapitalized young plants

So far, my empirical tests for the effect of competition on capital convergence have focused on MRPK convergence. The advantage of examining MRPK convergence is that any plant optimally converges to $MRPK^*_{it}$. Hence, the tests on MRPK convergence are generally valid for any type of plant, conditional on finding valid measures for competition and $MRPK^*_{it}$. Still, the empirical measurement of MRPK convergence is based on a certain set of assumptions, and necessitates the use of an autoregressive framework. To complement the evidence from MRPK convergence with evidence from more transparent estimation approaches, I now examine capital growth for young plants, which is a reduced-form object in the data.

7.1 Theoretical background for capital convergence of young plants

In my baseline model, capital wedges arise from the interaction of productivity volatility with financial constraints. In addition to productivity volatility, the birth of undercapitalized firms can also be a source of capital wedges. In fact, I formally show in Appendix D how a model with arrival of undercapitalized newborn firms in each period leads to isomorphic theoretical predictions for the effect of competition as in Proposition 1. Intuitively, when firms are born with suboptimally low levels of capital, then firms' optimizing behavior implies that firms grow their capital to its optimal level while they are young and financially constrained. In this setting, increased competition also reduces markups and thereby slows down internally financed capital growth for young plants.

7.2 Empirics for capital growth of young plants

Instead of empirically examining MRPK convergence, I can now analyze the impact of competition on capital growth of young plants, which requires less structural assumptions for its measurement. Throughout this empirical section, capital growth is measured as:[45]

$$g(k_{irst}) = k_{irst+1} - k_{irst}$$

Preliminary stylized fact A necessary condition to empirically find a negative effect of competition on capital growth, is that in the data, young plants are in the process of growing their capital in order to converge to their optimal capital level. The existing empirical literature provides extensive support for the stylized fact that young plants exhibit higher capital growth rates than older plants (see e.g. Evans (1987); Geurts and Van Biesebroeck (2014); Haltiwanger et al. (2013)). I demonstrate that its validity extends to the Indian manufacturing sector in Appendix Table A.2.

After validating the assumption that young plants are undercapitalized, I now examine the impact of competition on young plants' capital growth. As in the analysis of MRPK conver-

[45]Here, K_{irst+1} is the book value of assets at the end of the financial year. By measuring capital growth this way, it is not necessary to observe plants in previous years, which increases the sample size.

Table 4: Dereservation and capital growth for young plants

	Capital growth $g(k)_{irst}$			
	(1)	(2)	(3)	(4)
$Deres_{irst-1}$	-0.0374	-0.0443*	0.0111	0.00127
	(0.0238)	(0.0244)	(0.00985)	(0.0102)
$Deres_{irst} * [-\ln(age_{irst})]$	-0.0170**	-0.0138*		
	(0.00757)	(0.00789)		
$Deres_{irst-1} * 1(age_{irst} \leq 5)$			-0.0373**	-0.0735***
			(0.0172)	(0.0212)
$[-\ln(age_{irst})]$	0.0161***	0.0216***		
	(0.00295)	(0.00528)		
$1(age_{irst} \leq 5)$			0.0483***	0.0331***
			(0.00593)	(0.00751)
State-sector-year Fixed Effects	Yes	Yes	Yes	Yes
Plant Fixed Effects	No	Yes	No	Yes
Observations	107701	99316	109009	100663

P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the plant level.
The sample is restricted to all plants that were incumbent more than two years before their first product was dereserved. A plant is considered incumbent if it was producing at least 20% of its revenue share on a reserved product category in at least one year prior to dereservation.

gence, I first analyze the impact of the dereservation reform on young plants' capital growth, and then study the effect of the median markup.

7.2.1 Dereservation and capital growth of young plants

Estimation strategy To examine the effect of dereservation on capital growth for young plants, I run the following difference-in-difference specification.

$$\begin{aligned} g(k_{irst}) = \alpha_{irs} + \gamma_{rst} + \beta_1 young_{irst} + \beta_2 Deres_{irst-1} \\ + \beta_3 Deres_{irst-1} * young_{irst} + \beta_4 \ln age_{irst} + \varepsilon_{irst} \end{aligned} \tag{42}$$

Where I will consider two different measures for $young_{irst}$, namely $[-\ln(age_{irst})]$ and the indicator variable $1(age_{irst} \leq 5)$.[46] The prediction is that the increase in competition due to dereservation leads to slower capital growth for young plants, namely $\beta_3 < 0$.[47]

Estimation results Table 4 demonstrates that dereservation has a significantly negative impact on the capital growth for young plants. The magnitude of this impact is especially strong

[46] I will estimate specification (42) both with and without the plant-level fixed effects α_{irs}, as it is ambiguous whether these should be included or not. These fixed effects are the best way to control for unobserved plant-level characteristics. However, the theory predicts that capital growth ends once a plant reaches its optimal capital level, regardless of fixed plant-level characteristics.

[47] This specification allows estimation on a larger sample, as it does not require that lagged values are observed. Statistical power will prove not to be an issue, which is why I present results only for the sample of incumbents. Results for the full sample, which are comparable to those for the sample of incumbents, are available on request.

for very young plants. For instance, specification 4 estimates that for plants younger than 5 years old, dereservation leads to a reduction in the growth rate of 0.07 log points.[48]

7.2.2 Capital growth of young plants in the full sample

Estimation strategies In addition to the impact of dereservation on young plants' capital growth, it is useful to examine if the negative impact of competition on capital convergence for young plants generalizes to the full sample of Indian manufacturing plants. To this end, I again employ the median markup - $Median_{rst}[\ln(\mu_{irst-1})]$ - as an inverse measure of competition. Since the median markup is plausibly exogenous from the point of view of the individual plant, it allows me to use the full sample of plants. I update the regression analysis to the following specification, and predict that $\beta_2 > 0$.

$$g(k_{irst}) = \alpha_{irs} + \gamma_{rst} + \beta_1 young_{irst} + \beta_2 Median_{rst}[\ln(\mu_{irst-1})] * young_{irst} + \varepsilon_{irst} \tag{43}$$

I also examine the heterogeneous impact of competition across sectors with different levels of financial dependence. The prediction is that the impact of competition on capital growth for young plants is increasing with the degree of financial dependence ($\beta_3 > 0$).

$$\begin{aligned} g(k_{irst}) =& \alpha_{rst} + \beta_1 young_{irst} + \beta_2 Median_{rst}[\ln(\mu_{irst-1})] * young_{irst} \\ &+ \beta_3 Median_{rst}[\ln(\mu_{irst-1})] * young_{irst} * Fin\ Dep_s + \varepsilon_{irst} \end{aligned} \tag{44}$$

Results Also for this set of empirical tests, the estimation results are generally in line with the theoretical predictions (see Table 5). Capital growth for young plants increases with the median markup, and this effect is stronger in sectors with higher levels of financial dependence. These results are especially strong for the continuous measurement of age (columns 1,2,5,6), and less strong, with some insignificant results, when using the indicator variable for plants being younger than five years old.

[48] I use a slightly different definition of an incumbent plant in this section, compared to the previous analysis of dereservation. Previously, any plant producing a reserved product at any time prior to dereservation was considered incumbent. Here, I restrict the sample to incumbents where at least 20% of its revenue share was on a reserved product category at any time prior to dereservation.The results for the alternative definition of incumbent go in the same direction as the ones presented here, but are less strong.

Table 5: Competition and capital growth of young plants

	Capital growth $g(k)_{irst}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Median_{rst}[\ln(\mu_{irst})] * [-\ln(age)]$	0.00291***	0.00388**			0.00202	0.00472**		
	(0.00105)	(0.00154)			(0.00156)	(0.00237)		
$Median_{rst}[\ln(\mu_{irst})] * 1(age_{irst} \leq 5)$			0.00671***	-0.00364			0.00590*	-0.00271
			(0.00215)	(0.00307)			(0.00325)	(0.00474)
$Median_{rst}[\ln(\mu_{irst})] * [-\ln(age_{irst})] * Fin\ Dep_s$					0.00801**	0.000363		
					(0.00324)	(0.00487)		
$Median_{rst}[\ln(\mu_{irst})] * 1(age_{irst} \leq 5) * Fin\ Dep_s$							0.0111	-0.00533
							(0.00705)	(0.00998)
$-\ln(age_{irst})$	0.0131***	0.0175***			0.00957***	0.0125***		
	(0.00106)	(0.00239)			(0.00160)	(0.00325)		
$1(age_{irst} \leq 5)$			0.0585***	0.0210***			0.0494***	0.0149***
			(0.00207)	(0.00320)			(0.00304)	(0.00458)
$-\ln(age_{irst}) * Fin\ Dep_s$					0.0124***	0.0134**		
					(0.00263)	(0.00550)		
$1(age_{irst} \leq 5) * Fin\ Dep_s$							0.0266***	0.00822
							(0.00572)	(0.00875)
State-sector-year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	671592	552222	686466	565226	569555	463039	582655	474351

P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. standard errors are clustered at the plant level. The variable $Median_{rst-1}[\ln(\mu_{irst-1}]$ is demeaned by sector, and then measured in standard deviation units.

Columns 1-4 display estimation results for specification (43), while columns 5-8 show results for specification (44).

8 Conclusion

This paper examines how competition affects capital misallocation and markup misallocation. The starting point of the theory is a setting with firm-level productivity volatility, financial constraints, and imperfect competition. Competition then plays a dual role in affecting misallocation. On the one hand, both markup levels and markup dispersion tend to fall with competition, which unambiguously improves allocative efficiency in a setting without financial constraints. On the other hand, in a setting with financial constraints, a reduction in markups slows down capital accumulation, as the rate of self-financed investment shrinks. Thus, the positive impact of competition on steady-state misallocation is reduced by the presence of financial frictions. While the beneficial impact of competition is well known in the misallocation literature, the negative impact of competition in a setting with financial constraints was previously under examined.

A series of empirical tests confirm the model's predictions, in particular the prediction that competition slows down capital convergence. First, I explot India's 1997 natural variation in competition arising from India's 1997 dereservation reform, and show that this reform reduces markup-levels, and slows down capital convergence. Then, I demonstrate that the negative effect of competition is also present in the full sample of Indian manufacturing plants. Moreover, this effect is particularly pronounced in sectors with higher levels of financial dependence.

References

Aghion, P., Akcigit, U., and Howitt, P. (2013). What do we learn from schumpeterian growth theory? Technical report, National Bureau of Economic Research.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, pages 701–728.

Akcigit, U., Alp, H., and Peters, M. (2014). Lack of selection and imperfect managerial contracts: Firm dynamics in developing countries. *working paper, University of Pennsylvania*.

Allcott, H., Collard-Wexler, A., and O'Connell, S. D. (2014). How do electricity shortages affect productivity? Evidence from India. *NBER working paper*.

Asker, J., Collard-Wexler, A., and De Loecker, J. (2014). Dynamic inputs and resource (mis) allocation. *Journal of Political Economy*, 122(5):1013–1063.

Asturias, J., García-Santana, M., and Ramos, R. (2017). Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india.

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

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

Bénassy, J.-P. (1996). Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters*, 52(1):41–47.

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

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

Buera, F. J., Kaboski, J. P., and Shin, Y. (2015). Entrepreneurship and financial frictions: A macro-development perspective. Technical report, National Bureau of Economic Research.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *The American Economic Review*, 102(6):2437–2471.

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

Evans, D. S. (1987). Tests of alternative theories of firm growth. *The Journal of Political Economy*, pages 657–674.

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

Geurts, K. and Van Biesebroeck, J. (2014). Job creation, firm creation, and de novo entry. *CEPR Discussion Paper No. DP10118*.

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

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

Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). Who creates jobs? Small versus large versus young. *Review of Economics and Statistics*, 95(2):347–361.

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

Hsieh, C.-T. and Klenow, P. J. (2014). The Life Cycle of Plants in India and Mexico. *The Quarterly Journal of Economics*, 129(3):1035–1084.

Itskhoki, O. and Moll, B. (2015). Optimal development policies with financial frictions. *working paper, Princeton University*.

Jaimovich, N. (2007). Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations. *Journal of Economic Theory*, 137(1):300–325.

Kuntchev, V., Ramalho, R., Rodríguez-Meza, J., and Yang, J. S. (2014). What have we learned from the enterprise surveys regarding access to credit by SMEs? *World Bank Policy Research Working Paper*, (6670).

Macchiavello, R. and Morjaria, A. (2015). Competition and relational contracts: Evidence from rwanda's mills.

Martin, L. A., Nataraj, S., and Harrison, A. (2014). In with the big, out with the small: Removing small-scale reservations in india. *NBER working paper*.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

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

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: Journal of the Econometric Society*, pages 1417–1426.

Peters, M. (2013). Heterogeneous Mark-ups, Growth and Endogenous Misallocation. *mimeo, Yale University*.

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

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

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, 16(1):1–10.

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

Tewari, I. and Wilde, J. (2016). Product Scope and Productivity: Evidence from India's Product Reservation Policy. *working paper, University of South Florida*.

A Labor market equilibrium

A.1 Expressions for output and TFP

We can express each firm's capital as a share of aggregate capital. To that end, we rewrite capital demand for constrained and unconstrained firms as:

$$k_{it} = \mu_{it}^{\frac{1}{\eta-1}} a_{it}^{\frac{\eta}{1-\eta}} \frac{Q_t}{M} \left(\frac{P_t(1-\alpha)}{w_t}\right)^{\frac{\eta-\alpha\eta}{1-\eta}} \left(\frac{\alpha}{\omega_{it}}\right)^{\frac{1+\alpha\eta-\eta}{1-\eta}}$$

where $\omega_{it} = r_{it}$ if the firm is unconstrained, and $\omega_{it} > r_{it}$ otherwise. Writing k_{it} as a fraction of aggregate capital, we find:

$$k_{it} = \frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}} K_t$$

Similarly, for labor, starting from the labor demand equation $l_{it} = \left(\frac{(1-\alpha)}{\mu_{it}}\frac{P_t}{w_t}\left(\frac{Q_t}{M}\right)^{1-\eta} a_{it}^{\eta} k_{it}^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}$

$$l_{it} = \frac{\left(\frac{a_{it}^{\eta}k_{it}^{\alpha\eta}}{\mu_{it}}\right)^{\frac{1}{1+\alpha\eta-\eta}}}{\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}k_{it}^{\alpha\eta}}{\mu_{it}}\right)^{\frac{1}{1+\alpha\eta-\eta}}} L$$

Plugging in the value for k_{it}

$$l_{it} = \frac{\dfrac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{\alpha\eta}}\right)^{\frac{1}{1-\eta}}}{\left[\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}\right]^{\frac{\alpha\eta}{1+\alpha\eta-\eta}}}}{\sum_{i=1}^{M}\dfrac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{\alpha\eta}}\right)^{\frac{1}{1-\eta}}}{\left[\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}\right]^{\frac{\alpha\eta}{1+\alpha\eta-\eta}}}} L$$

The expressions for k_{it}, l_{it} can then be used to find an expression for the composite good: $Q_t = M^{1-\frac{1}{\eta}}\left[\sum_{i=1}^{M}(y_{it})^{\eta}\right]^{\frac{1}{\eta}} = M^{1-\frac{1}{\eta}}\left[\sum_{i=1}^{M}\left(a_{it}k_{it}^{\alpha}l_{it}^{1-\alpha}\right)^{\eta}\right]^{\frac{1}{\eta}}$.

$$Q_t = MK_t^{\alpha}L^{1-\alpha}\left[E_{it}\left(a_{it}^{\eta}\left(\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}\right)^{\alpha\eta}\left(\frac{\dfrac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{\alpha\eta}}\right)^{\frac{1}{1-\eta}}}{\left[\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}\right]^{\frac{\alpha\eta}{1+\alpha\eta-\eta}}}}{\sum_{i=1}^{M}\dfrac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{\alpha\eta}}\right)^{\frac{1}{1-\eta}}}{\left[\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}\right]^{\frac{\alpha\eta}{1+\alpha\eta-\eta}}}}\right)^{\eta-\alpha\eta}\right)\right]^{1/\eta}$$

Therefore:

$$Q_t = TFP_t K_t^{\alpha} L^{1-\alpha} \tag{45}$$

where

$$TFP_t \equiv M \left[E_{it} \left(a_{it}^{\eta} \left(\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}} \right)^{\alpha\eta} \left(\frac{\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{\alpha\eta}} \right)^{\frac{1}{1-\eta}}}{\left[\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}} \right]^{\frac{\alpha\eta}{1+\alpha\eta-\eta}}}}{\sum_{i=1}^{M} \frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{\alpha\eta}} \right)^{\frac{1}{1-\eta}}}{\left[\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}} \right]^{\frac{\alpha\eta}{1+\alpha\eta-\eta}}}} \right)^{\eta-\alpha\eta} \right) \right]^{1/\eta} \tag{46}$$

A.2 Labor market equilibrium

$$L = \sum_{i=1}^{M} \left(\frac{(1-\alpha)}{\mu_{it}} \frac{P_t}{w_t} \left(\frac{Q_t}{M} \right)^{1-\eta} a_{it}^{\eta} k_{it}^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}}$$

$$L = \left((1-\alpha) \frac{P_t}{w_t} \left(\frac{Q_t}{M} \right)^{1-\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} \sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}} k_{it}^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}}$$

$$L = \left((1-\alpha) \frac{P_t}{w_t} \left(\frac{TFP_t K_t^{\alpha} L^{1-\alpha}}{M} \right)^{1-\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} K_t^{\frac{\alpha\eta}{1+\alpha\eta-\eta}} \sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}} \left(\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}} \right)^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}}$$

$$L^{\frac{\alpha}{1+\alpha\eta-\eta}} = \left((1-\alpha) \frac{P_t}{w_t} \left(\frac{TFP_t}{M} \right)^{1-\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} K_t^{\frac{\alpha}{1+\alpha\eta-\eta}} \sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}} \left(\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}} \right)^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}}$$

$$\frac{P_t}{w_t} = \left(\frac{L}{K_t} \right)^{\alpha} \frac{1}{(1-\alpha)\left(\frac{TFP_t}{M} \right)^{1-\eta}} \left[\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}} \left(\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M} \left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}} \right)^{\frac{1}{1-\eta}}} \right)^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} \right]^{\eta-\alpha\eta-1}$$

So $\frac{P_t}{w_t}$ is decreasing in TFP and in $\left[\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}}\left(\frac{\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}{\sum_{i=1}^{M}\left(\frac{a_{it}^{\eta}}{\mu_{it}\omega_{it}^{1+\alpha\eta-\eta}}\right)^{\frac{1}{1-\eta}}}\right)^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}\right]$.

B Proof for Proposition 1

Proposition 1 has three components and the following appendix sections provide the proof for each of the three components. It will be convenient to first focus on the third component, namely the relation between capital wedges and competition.

B.1 Overview of the proof

First, I demonstrate what the sufficient conditions are for the proposition's statement for capital wedges, namely that $\forall \tau : \frac{d\omega_\tau}{dM} \geq 0$ and $(\Phi_\tau > 0) \implies \frac{d\omega_\tau}{dM}$. To demonstrate this, I start from equation (26), which for convenience is reiterated here:

$$\omega_\tau = G_\tau^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_H^\eta}{a_L^\eta} \frac{\mu_L}{\mu_\tau}\right]^{\frac{1}{1+\alpha\eta-\eta}} r_L$$

Therefore:

$$\begin{aligned}\frac{d\omega_\tau}{dM} =& -\frac{1-\eta}{1+\alpha\eta-\eta} G_\tau^{-\frac{1-\eta}{1+\alpha\eta-\eta}-1} \frac{dG_\tau}{dM} \left[\frac{a_H^\eta}{a_L^\eta} \frac{\mu_L}{\mu_\tau}\right]^{\frac{1}{1+\alpha\eta-\eta}} r_L \\ &+ \frac{1}{1+\alpha\eta-\eta} \left(\frac{\mu_L}{\mu_\tau}\right)^{\frac{1}{1+\alpha\eta-\eta}-1} \frac{d(\frac{\mu_L}{\mu_\tau})}{dM} G_\tau^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_H^\eta}{a_L^\eta}\right]^{\frac{1}{1+\alpha\eta-\eta}} r_L \\ &+ G_\tau^{-\frac{1-\eta}{1+\alpha\eta-\eta}} \left[\frac{a_H^\eta}{a_L^\eta} \frac{\mu_L}{\mu_\tau}\right]^{\frac{1}{1+\alpha\eta-\eta}} \frac{dr_L}{dM}\end{aligned}$$

In this preliminary version of the paper, I assume that $\frac{dr_L}{dM} = 0$. A sufficient condition for $\frac{d\omega_\tau}{dM} \geq 0$ and $(\Phi_\tau > 0) \implies \frac{d\omega_\tau}{dM}$ to hold, is then the following two conditions hold:

- $\forall \tau > 0 : ((\frac{dG_\tau}{dM} \leq 0) \wedge (\Phi_\tau > 0)) \implies \frac{dG_\tau}{dM} < 0$
- $\forall \tau > 0 : \frac{d(\frac{\mu_L}{\mu_\tau})}{dM} > 0$

Here is then the outline for the proof.

- First, I will derive an expression for capital growth, and show how it depends on μ_τ, μ_L.
- Then, for any $M' > M$, I will consider two cases: either $\mu_L' \geq \mu_L$ or $\mu_L' < \mu_L$. I demonstrate that $\mu_L' \geq \mu_L$ results in a contradiction and therefore $\mu_L' < \mu_L$ holds. Intuitively, $\mu_L' \geq \mu_L$ leads to a contradiction, because it implies a higher market share for a_L-type firms, while at the same time increasing G_τ and thereby inducing higher market shares for a_H-type firms as well. Increasing market shares for both a_L, a_H-type firms then contradicts with the average market share decreasing with M.
- I then show that $\mu_L' < \mu_L$ implies that
  - $\forall \tau > 0 : G_\tau' \leq G_\tau \wedge ((\Phi_\tau > 0) \implies G_\tau' < G_\tau)$
  - $\forall \tau > 0 : \mu_\tau' \leq \mu_\tau$
  - $\forall \tau > 0 : \frac{\mu_L'}{\mu_\tau'} > \frac{\mu_L}{\mu_\tau}$

  which concludes the proof. This pattern for markups and capital growth is intuitive: increased M lowers markups for all types of firms, which at the same time reduces capital growth for financially constrained firms. The theoretical challenge lies in demonstrating that this is the only possible pattern for markups and capital growth.

B.2 Expression for capital growth

I consider capital growth for all firms with $\tau \geq 0$. This type of firms are only heteregenous across different bins τ, and are perfectly homogeneous within a bin τ. At the same time, as explained in the paper, capital k_τ for these firms is predetermined, and their productivity a_τ is exogenous: $a_0 = a_L, \forall \tau > 0 : a_\tau = a_H$.

Financially constrained firms, i.e. firms with $\Phi_\tau > 0$, invest all their retained earnings into capital investment. Therefore, for a financially constrained firm in bin τ, capital growth $g(k_\tau) = \frac{(\mu_\tau - \frac{AC_\tau}{MC_\tau})y_\tau MC_\tau}{k_\tau} - \delta$, where AC_τ is average cost and MC_τ is marginal cost.

The firm's total costs, for any quantity $\bar{y}_\tau$ are $TC(\bar{y}_\tau) = \frac{w}{P}L(\bar{y}_\tau)$. Here, since a_τ, and k_τ are exogenous and predetermined, respectively, setting $\bar{y}_\tau$ directly implies setting $\bar{l}_\tau$ since $\bar{y}_\tau = a_H k_\tau^\alpha \bar{l}_\tau^{1-\alpha}$. This means that $L(\bar{y}_\tau) = \left(\frac{\bar{y}_\tau}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}}$, such that $TC(\bar{y}_\tau) = \frac{w}{P}\left(\frac{\bar{y}_\tau}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}}$

Therefore:

$$MC_\tau(\bar{y}_\tau) = \frac{\partial TC(\bar{y}_\tau)}{\partial \bar{y}_\tau} = \frac{w}{(1-\alpha)P}\left(\frac{\bar{y}_\tau^\alpha}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}}$$

$$AC_\tau(\bar{y}_\tau) = \frac{w}{P}\frac{1}{\bar{y}_\tau}\left(\frac{\bar{y}_\tau}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}} = \frac{w}{P}\left(\frac{\bar{y}_\tau^\alpha}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}}$$

Which implies that

$$\frac{AC_\tau(\bar{y}_\tau)}{MC_\tau(\bar{y}_\tau)} = (1-\alpha)$$

Capital growth expression Start with derivation of profits, where $\bar{\mu}_\tau$ is determined by choosing $\bar{y}_\tau$ and setting the price given the demand function.

$$\pi_\tau = \left(\bar{\mu}_\tau - \frac{AC_\tau}{MC_\tau}\right)\bar{y}_\tau * MC_\tau = (\bar{\mu}_\tau - (1-\alpha))\frac{w}{(1-\alpha)P}\left(\frac{\bar{y}_\tau^\alpha}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}}\bar{y}_\tau$$

$$\pi_\tau = (\bar{\mu}_\tau - (1-\alpha))\frac{w}{P(1-\alpha)}\left(\frac{\bar{y}_\tau}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}}$$

Hence,

$$\frac{\pi_\tau}{k_\tau} = (\bar{\mu}_\tau - (1-\alpha))\frac{w}{P(1-\alpha)k_\tau}\left(\frac{\bar{y}_\tau}{a_H k_\tau^\alpha}\right)^{\frac{1}{1-\alpha}} = [\bar{\mu}_\tau - (1-\alpha)]\frac{w}{P(1-\alpha)}\left(\frac{\bar{y}_\tau}{a_H k_\tau}\right)^{\frac{1}{1-\alpha}}$$

and since $\bar{y}_\tau = a_\tau k_\tau^\alpha \bar{l}_\tau^{1-\alpha}$

$$\frac{\pi_\tau}{k_\tau} = (\bar{\mu}_\tau - (1-\alpha))\frac{w}{P(1-\alpha)}\left(\frac{a_\tau \bar{l}_\tau^{1-\alpha}}{a_\tau k_\tau^{1-\alpha}}\right)^{\frac{1}{1-\alpha}} = [\bar{\mu}_\tau - (1-\alpha)]\frac{w}{P(1-\alpha)}\frac{\bar{l}_\tau}{k_\tau}$$

Since all profits are invested in capital growth, we have for $\Phi_\tau > 0$:

$$\frac{k_{\tau+1}(\bar{l}_\tau) - k_\tau}{k_\tau} = \frac{\pi_\tau}{k_\tau} - \delta = [\bar{\mu}_\tau - (1-\alpha)]\frac{w}{P(1-\alpha)}\frac{\bar{l}_\tau}{k_\tau} - \delta$$

Given the expression for $\frac{\pi_\tau}{k_\tau}$, we need to determine μ_τ, $\frac{\bar{l}_\tau}{k_\tau}$. These variables are outcomes of the optimization problem, where optimal labor l_τ is from equation (18) in the paper, while for capital $k_\tau = G_\tau k_L$. Finally, the markup is also optimally determined as μ_τ, defined in equation (24).

Remember:

$$l_\tau = \left(\frac{(1-\alpha)}{\mu_\tau}\frac{P}{w}\left(\frac{Q}{M}\right)^{1-\eta} a_\tau^\eta k_\tau^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}$$

$$k_L^* = \left(\frac{a_L^\eta}{\mu_L}\right)^{\frac{1}{1-\eta}}\left(\frac{\alpha}{r_L}\right)^{\frac{1+\alpha\eta-\eta}{1-\eta}}\left(\frac{P(1-\alpha)}{w}\right)^{\frac{\eta-\alpha\eta}{1-\eta}}\frac{Q}{M}$$

therefore

$$\frac{l_\tau^*}{G_\tau k_L^*} = \frac{P(1-\alpha)}{G_\tau w}\left(\frac{a_\tau}{a_L}\right)^{\frac{\eta}{1+\alpha\eta-\eta}}\frac{r_L}{\alpha}\left(\frac{\mu_L}{\mu_\tau}\right)^{\frac{1}{1+\alpha\eta-\eta}}$$

which implies that for $g(k)_\tau \equiv \frac{k_{\tau+1}(l_\tau)-k_\tau}{k_\tau}$

$$\forall\tau \text{ where } \Phi_\tau > 0 : g(k)_\tau = \frac{1}{G_\tau}\left[\mu_\tau - (1-\alpha)\right]\left(\frac{a_\tau}{a_L}\right)^{\frac{\eta}{1+\alpha\eta-\eta}}\frac{r_L}{\alpha}\left(\frac{\mu_L}{\mu_\tau}\right)^{\frac{1}{1+\alpha\eta-\eta}} - \delta$$

or

$$\forall\tau \text{ with } \Phi_\tau > 0 : \ln(G_\tau(g(k)_\tau+\delta)) = \ln\left[\mu_\tau-(1-\alpha)\right] + \ln\left(\frac{a_\tau}{a_L}\right)^{\frac{\eta}{1+\alpha\eta-\eta}}\frac{r_L}{\alpha} - \frac{\ln\left(\frac{\mu_\tau}{\mu_L}\right)}{1+\alpha\eta-\eta} \tag{47}$$

Derivative with respect to M Assuming r_L is constant, and only considering τ where $\Phi_\tau > 0$ we find that

$$\frac{\partial\ln(G_\tau(g(k)_\tau+\delta))}{\partial M} = \frac{\frac{\partial\mu_\tau}{\partial M}}{[\mu_\tau-(1-\alpha)]} - \frac{1}{1+\alpha\eta-\eta}\frac{\mu_L}{\mu_\tau}\left[\frac{\partial\mu_\tau}{\partial M}\frac{1}{\mu_L} - \frac{\mu_\tau}{\mu_L^2}\frac{\partial\mu_L}{\partial M}\right]$$

Rearranging:

$$\frac{\partial\ln(G_\tau(g(k)_\tau+\delta))}{\partial M} = \frac{\partial\mu_\tau}{\partial M}\left(\frac{1}{[\mu_\tau-(1-\alpha)]} - \frac{1}{1+\alpha\eta-\eta}\frac{1}{\mu_\tau}\left[1-\frac{\mu_\tau}{\mu_L}\frac{\frac{\partial\mu_L}{\partial M}}{\frac{\partial\mu_\tau}{\partial M}}\right]\right) \tag{48}$$

Note that a sufficient condition for $sign(\frac{\partial\ln(G_\tau(g(k)_\tau+\delta))}{\partial M}) = sign(\frac{\partial\mu_\tau}{\partial M})$, is that $sign(\frac{\partial\mu_\tau}{\partial M}) = sign(\frac{\partial\mu_L}{\partial M})$. This is because

$$\left(\frac{1}{[\mu_\tau-(1-\alpha)]} - \frac{1}{1+\alpha\eta-\eta}\frac{1}{\mu_\tau}\left[1-\frac{\mu_\tau}{\mu_L}\frac{\frac{\partial\mu_L}{\partial M}}{\frac{\partial\mu_\tau}{\partial M}}\right] > 0\right) \iff \left(1 > \frac{\left[1-\frac{(1-\alpha)}{\mu_\tau}\right]}{1-\eta(1-\alpha)}\left[1-\frac{\mu_\tau}{\mu_L}\frac{\frac{\partial\mu_L}{\partial M}}{\frac{\partial\mu_\tau}{\partial M}}\right]\right)$$

and $\left(sign(\frac{\partial\mu_\tau}{\partial M}) = sign(\frac{\partial\mu_L}{\partial M})\right) \implies \left(0 > -\frac{\mu_\tau}{\mu_L}\frac{\frac{\partial\mu_L}{\partial M}}{\frac{\partial\mu_\tau}{\partial M}}\right)$. Hence, a key step in the remainder of the proof will be demonstrating that $\left(sign(\frac{\partial\mu_\tau}{\partial M}) = sign(\frac{\partial\mu_L}{\partial M})\right)$ holds globally.

B.3 Impact of competition on distribution of markups and capital growth

Given equation (48), I will now examine the level of markups and capital growth, across any two different levels for the number of firms in the economy, namely $M' > M$, where I denote with a prime the values under M'. Specifically, I will examine two cases. First, $\mu'_L \geq \mu_L$ and second $\mu'_L < \mu_L$. The first case will result in a contradiction, so its opposite - the second case - must be true. The analysis in the second case will then characterize the path of markups and capital growth across different M. For the analysis, it will be useful to define $\mathcal{G}_\tau \equiv \frac{y_\tau^\eta}{y_L^\eta}$.

B.3.1 Case 1: $\mu'_L \geq \mu_L$

This case will result in a contradiction, and therefore its opposite must be true. The proof proceeds by induction.

Step 1 Consider $\tau = 0$, where productivity is a_L and $\mu_0 = \mu_L$, but the firm learns it will have productivity a_H in $\tau = 1$. In this period, if $\Phi_0 > 0$, capital growth is

$$g(k)_0 = [\mu_L - (1-\alpha)] \frac{r_L}{\alpha} \tag{49}$$

Therefore $g(k)'_0 \geq g(k)_0$ since $\mu'_L \geq \mu_L$ and the other variables are constant.

Inductive step For the inductive step, I show first that for any period $\tau > 0$ with $G'_\tau \geq G_\tau$:

$$\left((G'_\tau \geq G_\tau) \wedge (\mu'_L \geq \mu_L)\right) \implies (\mu'_\tau \geq \mu_\tau)$$

To show this, notice that $(\mathcal{G}'_\tau \geq \mathcal{G}_\tau)/(\mathcal{G}'_\tau < \mathcal{G}_\tau)$, and in both cases, I show that $(\mu'_\tau \geq \mu_\tau)$ holds

- Case (i): $((\mathcal{G}'_\tau \geq \mathcal{G}_\tau) \wedge (G'_\tau \geq G_\tau) \wedge (\mu'_L \geq \mu_L)) \implies (\mu'_\tau \geq \mu_\tau)$. This follows from $((\mathcal{G}'_\tau \geq \mathcal{G}_\tau) \wedge (\mu'_L \geq \mu_L)) \implies (\mu'_\tau \geq \mu_\tau)$.
  - From equations (9), (10), it is clear that μ_L is monotonically increasing in $\frac{y_L^\eta}{\sum y_{it}^\eta}$. Therefore, $(\mu'_L \geq \mu_L) \iff \left(\frac{y_L'^\eta}{\sum y_{it}'^\eta} \geq \frac{y_L^\eta}{\sum y_{it}^\eta}\right)$. Then, $(\mathcal{G}'_\tau \geq \mathcal{G}_\tau) \wedge \left(\frac{y_L'^\eta}{\sum y_{it}'^\eta} \geq \frac{y_L^\eta}{\sum y_{it}^\eta}\right) \implies \left(\frac{\mathcal{G}'_\tau y_L'^\eta}{\sum y_{it}'^\eta} \geq \frac{\mathcal{G}_\tau y_L^\eta}{\sum y_{it}^\eta} \wedge \mu'_\tau \geq \mu_\tau\right)$
- Case (ii): $((\mathcal{G}'_\tau < \mathcal{G}_\tau) \wedge (G'_\tau \geq G_\tau) \wedge (\mu'_L \geq \mu_L)) \implies (\mu'_\tau \geq \mu_\tau)$.
  - Note that $\mathcal{G}_\tau \equiv \frac{y_\tau^\eta}{y_L^\eta} = \frac{(a_H G_\tau^\alpha l_\tau^{1-\alpha})^\eta}{(a_L l_L^{1-\alpha})^\eta}$. Therefore $(\mathcal{G}'_\tau < \mathcal{G}_\tau) \implies \left((G'_\tau < G_\tau) \vee (\frac{l'_\tau}{l'_L} < \frac{l_\tau}{l_L})\right)$ such that $\left((G'_\tau \geq G_\tau) \wedge (\frac{l'_\tau}{l'_L} \geq \frac{l_\tau}{l_L})\right) \implies (\mathcal{G}'_\tau \geq \mathcal{G}_\tau)$
  - Note that $\frac{l_\tau}{l_L} = \left(\frac{\mu_L}{\mu_\tau} \frac{a_H}{a_L} G_\tau^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}$. Therefore, $\left((\frac{l'_\tau}{l'_L} < \frac{l_\tau}{l_L}) \wedge (G'_\tau \geq G_\tau)\right) \implies \frac{\mu'_L}{\mu'_\tau} < \frac{\mu_L}{\mu_\tau}$
  - In this case, $(\mathcal{G}'_\tau < \mathcal{G}_\tau) \wedge (G'_\tau \geq G_\tau) \implies \left(\frac{l'_\tau}{l'_L} < \frac{l_\tau}{l_L}\right)$. However, $\left(\frac{l'_\tau}{l'_L} < \frac{l_\tau}{l_L}\right) \implies \frac{\mu'_L}{\mu'_\tau} < \frac{\mu_L}{\mu_\tau}$. Therefore, $((\mathcal{G}'_\tau < \mathcal{G}_\tau) \wedge (G'_\tau \geq G_\tau) \wedge (\mu'_L \geq \mu_L)) \implies (\mu'_\tau > \mu_\tau)$.
- Therefore, $((G'_\tau \geq G_\tau) \wedge (\mu'_L \geq \mu_L)) \implies (\mu'_\tau \geq \mu_\tau)$,

Given equation (48), $((\Phi_\tau > 0) \wedge (\mu'_\tau > \mu_\tau) \wedge (\mu'_L \geq \mu_L) \wedge (G'_\tau \geq G_\tau)) \implies G'_{\tau+1} \geq G_{\tau+1}$. This completes the inductive step.

Final step In case $\Phi_0 > 0$, then $g(k)'_0 \geq g(k)_0$. Hence, $(G'_1 \geq G_1) \implies (\mu'_1 \geq \mu_1)$. Therefore, this proof by induction implies that $(\mu'_L \geq \mu_L) \implies (\forall \tau > 0 \wedge \Phi_\tau > 0 : (\mu'_\tau \geq \mu_\tau)$. At the same time, $((M' > M) \wedge (\mu'_L > \mu_L)) \implies \exists \tau > 0 : (\mu'_\tau < \mu_\tau)$, which will yield a contradiction.

- To see that $((M' > M) \wedge (\mu'_L > \mu_L)) \implies \exists \tau : (\mu'_\tau < \mu_\tau)$, note that equations (9) and (10) for the markup and the demand elasticity entail that for any firm i: $(\mu'_{it} \geq \mu_{it}) \iff \left(\frac{y'^\eta_{it}}{\sum_i y'^\eta_{it}} \geq \frac{y^\eta_{it}}{\sum_i y^\eta_{it}}\right)$. Suppose for $M' > M$, we have $((\mu'_L \geq \mu_L) \wedge (\forall \tau > 0 : \mu'_\tau \geq \mu_\tau)) \implies \left((\frac{y'^\eta_L}{\sum_i y'^\eta_{it}} \geq \frac{y^\eta_L}{\sum_i y^\eta_{it}}) \wedge (\frac{y'^\eta_\tau}{\sum_i y'^\eta_{it}} \geq \frac{y^\eta_\tau}{\sum_i y^\eta_{it}})\right)$. Then,

$$\left((\frac{y'^\eta_L}{\sum_i y'^\eta_{it}} \geq \frac{y^\eta_L}{\sum_i y^\eta_{it}}) \wedge (\frac{y'^\eta_\tau}{\sum_i y'^\eta_{it}} \geq \frac{y^\eta_\tau}{\sum_i y^\eta_{it}})\right)$$
$$\implies \left(\frac{M'[Prob(a_{it} = a_L)\frac{y'^\eta_L}{\sum_i y'^\eta_{it}} + \sum_{\tau=1}^{\infty} Prob((a_{it} = a_H)\&(t = \tau))\frac{y'^\eta_\tau}{\sum_i y'^\eta_{it}}]}{M[Prob(a_{it} = a_L)\frac{y^\eta_L}{\sum_i y^\eta_{it}} + \sum_{\tau=1}^{\infty} Prob((a_{it} = a_H)\&(t = \tau))\frac{y^\eta_\tau}{\sum_i y^\eta_{it}}]} > 1\right)$$

  Which is a contradiction since both the denominator and the numerator in the ratio after the implication are equal to 1. This is because $\sum_i y^\eta_{it} = M[Prob(a_{it} = a_L)y^\eta_L + \sum_{\tau=1}^{\infty} Prob((a_{it} = a_H)\&(t = \tau))y^\eta_\tau]$. Since $M' > M \wedge ((\mu'_L \geq \mu_L) \wedge (\forall \tau > 0 : \mu'_\tau \geq \mu_\tau))$ entails a contradiction, $\exists \tau > 0 : (\mu'_\tau < \mu_\tau)$, under the continued assumption that $\mu'_L \geq \mu_L$.

- $(\exists \tau > 0 : (\mu'_\tau < \mu_\tau)) \implies ((\exists \tau > 0 : \Phi_\tau > 0 \wedge (\mu'_\tau < \mu_\tau)) \vee (\exists \tau > 0 : \Phi_\tau = 0 \wedge (\mu'_\tau < \mu_\tau)))$, but both cases result in a contradiction.
  - Case a: $(\exists \tau > 0 : \Phi_\tau > 0 \wedge (\mu'_\tau < \mu_\tau))$. This does not hold, since the proof by induction implies $(\mu'_L \geq \mu_L) \implies (\forall \tau > 0 \wedge \Phi_\tau > 0 : (\mu'_\tau \geq \mu_\tau)$.
  - Case b is equivalent to $(\mu'_H < \mu_H)$. We know that $(\mu'_L > \mu_L) \wedge (\mu'_H < \mu_H) \implies (\mathcal{G}'_H < \mathcal{G}_H)$, where $\mathcal{G}_H = \frac{(a_H G^\alpha_H l^{1-\alpha}_H)^\eta}{a_L l^{1-\alpha}_L)^\eta}$. Hence, $(\mathcal{G}'_H < \mathcal{G}_H) \implies \left((G'_H < G_H) \vee (\frac{l'_H}{l'_L} < \frac{l_H}{l_L})\right)$. There are then again two cases, both of which result in a contradiction:
    * Case b1: since $G_H = (\frac{a_H}{a_L}\frac{\mu_L}{\mu_H})^{1/(1-\eta)}$, $(G'_H < G_H) \implies (\frac{\mu'_L}{\mu'_H} < \frac{\mu_L}{\mu_H})$. However, $(\mu'_L > \mu) \wedge (\frac{\mu'_L}{\mu'_H} < \frac{\mu_L}{\mu_H}) \implies (\mu'_H > \mu_H)$, which contradicts the supposition that $(\mu'_H < \mu_H)$
    * Case b2: Since $\frac{l_\tau}{l_L} = \left(\frac{\mu_L}{\mu_\tau}\frac{a_H}{a_L}G^{\alpha\eta}_\tau\right)^{\frac{1}{1+\alpha\eta-\eta}}$, $(\frac{l'_H}{l'_L} < \frac{l_H}{l_L}) \wedge (G'_H \geq G_H) \implies (\frac{\mu'_L}{\mu'_H} < \frac{\mu_L}{\mu_H})$, which again results in a contradiction

Since the supposition that $\mu'_L \geq \mu_L$ entails a contradiction, its opposite must be true: $\mu'_L < \mu_L$

B.3.2 Case 2: $\mu'_L < \mu_L$

Step 1 Consider $\tau = 0$, from equation (49), it is clear that

$$((\mu'_L < \mu_L) \wedge (\Phi_0 > 0)) \implies \left((g(k)'_0 < g(k)_0) \wedge (G'_1 < G_1)\right)$$

Inductive step For the inductive step, I show first that for any period $\tau > 0$:

$$(G'_\tau < G_\tau) \wedge (\mu'_L < \mu_L) \implies (\mu'_\tau < \mu_\tau)$$

To prove that $(G'_\tau < G_\tau) \wedge (\mu'_L < \mu_L) \implies (\mu'_\tau < \mu_\tau)$, consider two cases:

- Case (i): $(G'_\tau < G_\tau) \wedge (\mu'_L < \mu_L) \wedge (\mathcal{G}'_\tau \leq \mathcal{G}_\tau) \implies (\mu'_\tau < \mu_\tau)$.
- Case (ii): $(G'_\tau < G_\tau) \wedge (\mu'_L < \mu_L) \wedge (\mathcal{G}'_\tau > \mathcal{G}_\tau) \implies (\mu'_\tau < \mu_\tau)$. This is because given $\mathcal{G}_\tau = \frac{(a_H G_\tau l_\tau)^\eta}{(a_L l_L)^\eta}$ and $\frac{l_\tau}{l_L} = \left(\frac{\mu_L}{\mu_\tau}\frac{a_H}{a_L}G_\tau^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}$; $((G'_\tau < G_\tau) \wedge (\mathcal{G}'_\tau > \mathcal{G}_\tau)) \implies (\frac{l'_1}{l'_L} > \frac{l_1}{l_L})$ and $\left((\mu'_L < \mu_L) \wedge (\frac{l'_1}{l'_L} > \frac{l_1}{l_L}) \wedge (G'_\tau < G_\tau)\right) \implies \left((\frac{\mu'_L}{\mu'_\tau} > \frac{\mu_L}{\mu_\tau}) \iff (1 > \frac{\mu'_L}{\mu_L} > \frac{\mu'_\tau}{\mu_\tau})\right)$
- Therefore, $(G'_\tau < G_\tau) \wedge (\mu'_L < \mu_L) \implies (\mu'_\tau < \mu_\tau)$

Given equation (48), $((\Phi_\tau > 0) \wedge (\mu'_\tau < \mu_\tau) \wedge (\mu'_L < \mu_L) \wedge (G'_\tau < G_\tau)) \implies G'_{\tau+1} < G_{\tau+1}$. This completes the inductive step, which applies for any $\tau > 0$ with $\Phi_\tau > 0$.

Final step We know that when $\Phi_0 > 0$, $g(k)'_0 < g(k)_0$. Hence, $\Phi_0 > 0 \implies [(G'_1 < G_1) \implies (\mu'_1 < \mu_1)]$. Therefore, this proof by induction implies that

$$(\mu'_L < \mu_L) \implies \left[(\Phi_\tau > 0) \implies ((\mu'_\tau < \mu_\tau) \wedge (G'_\tau < G_\tau))\right]$$

Result for μ_H, G_H How do μ_H, G_H evolve with M? There are two cases: $\mathcal{G}'_H \leq \mathcal{G}_H$ or $\mathcal{G}'_H > \mathcal{G}_H$.

- Case a: $(\mathcal{G}'_H \leq \mathcal{G}_H) \implies (\mu'_H < \mu_H)$. Why? We know that $\mu'_L < \mu_L \iff (\frac{y'^\eta_L}{\sum_i y'^\eta_{it}} < \frac{y^\eta_L}{\sum_i y^\eta_{it}})$. Hence, $((\mu'_L < \mu_L) \wedge (\mathcal{G}'_H \leq \mathcal{G}_H)) \implies \left((\mathcal{G}'_H \frac{y'^\eta_L}{\sum_i y'^\eta_{it}} < \frac{\mathcal{G}_H y^\eta_L}{\sum_i y^\eta_{it}}) \iff (\mu'_H < \mu_H)\right)$
- Case b: $(\mathcal{G}'_H > \mathcal{G}_H)$.
  - Note that $\mathcal{G}_\tau \equiv \frac{y^\eta_H}{y^\eta_L} = \frac{(a_H G_\tau l_H)^\eta}{(a_L l_L^{1-\alpha})^\eta}$. Hence, $(\mathcal{G}'_H > \mathcal{G}_H) \iff (\frac{G'^\alpha_H l'^{1-\alpha}_H}{l'_L} < \frac{G^\alpha_H l^{1-\alpha}_H}{l_L})$. Therefore $(\mathcal{G}'_H > \mathcal{G}_H) \implies \left((G'_H > G_H) \vee (\frac{l'_H}{l'_L} > \frac{l_H}{l_L})\right)$. There are then again two cases
  - Case b1: suppose $(G'_H > G_H)$. Since $G_H = (\frac{a_H}{a_L}\frac{\mu_L}{\mu_H})^{1/(1-\eta)}$, $(G'_H > G_H) \implies \left((\frac{\mu'_L}{\mu'_H} > \frac{\mu_L}{\mu_H}) \iff (\frac{\mu'_L}{\mu_L} > \frac{\mu'_H}{\mu_H})\right)$. Therefore, $(G'_H > G_H) \wedge (\mu'_L < \mu_L) \implies$
  - Case b2: suppose $(\frac{l'_H}{l'_L} > \frac{l_H}{l_L}) \wedge (G'_H \leq G_H)$. Note that $\frac{l_H}{l_L} = \left(\frac{\mu_L}{\mu_H}\frac{a_H}{a_L}G_H^{\alpha\eta}\right)^{\frac{1}{1+\alpha\eta-\eta}}$. Therefore, $\left((\frac{l'_H}{l'_L} > \frac{l_H}{l_L}) \wedge (G'_H \leq G_H)\right) \implies (\frac{\mu'_L}{\mu'_H} > \frac{\mu_L}{\mu_H} \iff \frac{\mu'_L}{\mu_L} > \frac{\mu'_H}{\mu_H})$. Since $1 > \frac{\mu'_L}{\mu_L}$, we find that $(\frac{l'_H}{l'_L} > \frac{l_H}{l_L}) \wedge (G'_H \leq G_H) \implies (\mu'_H < \mu_H)$.
  - Therefore, we find that $(\mu'_H < \mu_H)$ and hence

$$(\mu'_L < \mu_L) \implies [\forall \tau > 0 : ((\mu'_\tau < \mu_\tau))]$$

B.4 Relative markups and M

From the previous subsection, I know that

$$(\mu'_L < \mu_L) \wedge (\forall \tau > 0 : \mu'_\tau < \mu_\tau) \wedge \left((\Phi_\tau > 0) \implies (G'_\tau < G_\tau)\right)$$

As is already clear from the inductive step in subsection B.3.2, there are two cases: either $\mathcal{G}'_\tau \leq \mathcal{G}_\tau$ or $\mathcal{G}'_\tau > \mathcal{G}_\tau$. I now demonstrate that in both cases, $\frac{\mu'_L}{\mu'_\tau} > \frac{\mu_L}{\mu_\tau}$.

- In case $\mathcal{G}'_\tau > \mathcal{G}_\tau$, then case (ii) in subsection B.3.2 demonstrates that

$$((G'_\tau < G_\tau) \wedge (\mu'_L < \mu_L) \wedge (\mathcal{G}'_\tau > \mathcal{G}_\tau)) \implies \left(\frac{\mu'_L}{\mu'_\tau} > \frac{\mu_L}{\mu_\tau}\right)$$

- In case $\mathcal{G}'_\tau \leq \mathcal{G}_\tau$, then start from the expression for relative markups, derived from equation (9):

$$\frac{\mu_L}{\mu_\tau} = \frac{\frac{1-\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\eta\left(1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}\right)}}{\frac{1-\eta\frac{y_\tau^\eta}{\sum_i y_{it}^\eta}}{\eta\left(1-\frac{y_\tau^\eta}{\sum_i y_{it}^\eta}\right)}} = \frac{1-\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\left(1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}\right)} \frac{\left(1-\frac{y_\tau^\eta}{\sum_i y_{it}^\eta}\right)}{1-\eta\frac{y_\tau^\eta}{\sum_i y_{it}^\eta}}$$

First, define $\frac{y_\tau^\eta}{y_L^\eta} \equiv \mathcal{G}_\tau$, such that:

$$\frac{\mu_L}{\mu_\tau} = \frac{\left(1-\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}\right)}{\left(1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}\right)} \frac{\left(1-\frac{\mathcal{G}_\tau y_L^\eta}{\sum_i y_{it}^\eta}\right)}{\left(1-\eta\frac{\mathcal{G}_\tau y_L^\eta}{\sum_i y_{it}^\eta}\right)} = \frac{1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}(\mathcal{G}_\tau+\eta)+\eta\mathcal{G}_\tau(\frac{y_L^\eta}{\sum_i y_{it}^\eta})^2}{1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}(\mathcal{G}_\tau\eta+1)+\eta\mathcal{G}_\tau(\frac{y_L^\eta}{\sum_i y_{it}^\eta})^2}$$

Define: $Num \equiv 1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}(\mathcal{G}_\tau+\eta)+\eta\mathcal{G}_\tau(\frac{y_L^\eta}{\sum_i y_{it}^\eta})^2$ and $Denom \equiv 1-\frac{y_L^\eta}{\sum_i y_{it}^\eta}(\mathcal{G}_\tau\eta+1)+\eta\mathcal{G}_\tau(\frac{y_L^\eta}{\sum_i y_{it}^\eta})^2$ and find that

$$\begin{aligned}
&\frac{\partial\frac{\mu_L}{\mu_\tau}}{\partial M} * Denom^2 = \\
&\left[-\frac{\partial\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\partial M}(\mathcal{G}_\tau+\eta)-\frac{y_L^\eta}{\sum_i y_{it}^\eta}\frac{\partial\mathcal{G}_\tau}{\partial M}+2\mathcal{G}_\tau\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}\frac{\partial\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\partial M}+\eta(\frac{y_L^\eta}{\sum_i y_{it}^\eta})^2\frac{\partial\mathcal{G}_\tau}{\partial M}\right] * Denom \\
&- Num\left[-\frac{\partial\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\partial M}(\mathcal{G}_\tau\eta+1)-\frac{y_L^\eta}{\sum_i y_{it}^\eta}\eta\frac{\partial\mathcal{G}_\tau}{\partial M}+2\eta\mathcal{G}_\tau\frac{\partial\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\partial M}+\eta(\frac{y_L^\eta}{\sum_i y_{it}^\eta})^2\frac{\partial\mathcal{G}_\tau}{\partial M}\right]
\end{aligned}$$

Rearranging the RHS:

$$\begin{aligned}
&\frac{\partial\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\partial M}\left[Num(\mathcal{G}_\tau\eta+1)-Denom(\mathcal{G}_\tau+\eta)+2\frac{y_L^\eta}{\sum_i y_{it}^\eta}\mathcal{G}_\tau\eta(Denom-Num)\right] \\
&+\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}\frac{\partial\mathcal{G}_\tau}{\partial M}\left[(Num-\frac{Denom}{\eta})+\frac{y_L^\eta}{\sum_i y_{it}^\eta}(Denom-Num)\right]
\end{aligned}$$

Note that $\left[(Num-\frac{Denom}{\eta})+\frac{y_L^\eta}{\sum_i y_{it}^\eta}(Denom-Num)\right] < 0$ for any $\frac{y_L^\eta}{\sum_i y_{it}^\eta} < 1$ since $\eta < 1$ and $Denom > Num$. Since $\frac{\partial\frac{y_L^\eta}{\sum_i y_{it}^\eta}}{\partial M} < 0$ because $\mu'_L < \mu_L$ and because I assume that $\frac{\partial\mathcal{G}_\tau}{\partial M} < 0$, the following condition is sufficient for $\frac{\partial\frac{\mu_L}{\mu_\tau}}{\partial M} > 0$ to hold:

$$\left[Num(\mathcal{G}_\tau\eta+1)-Denom(\mathcal{G}_\tau+\eta)+2\mathcal{G}_\tau\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}(Denom-Num)\right] < 0$$

or

$$2\mathcal{G}_\tau\eta\frac{y_L^\eta}{\sum_i y_{it}^\eta}(Denom-Num) < Denom(\mathcal{G}_\tau+\eta)-Num(\mathcal{G}_\tau\eta+1)$$

$$\frac{y_L^\eta}{\sum_i y_{it}^\eta} < \frac{Denom(\mathcal{G}_\tau + \eta) - Num(\mathcal{G}_\tau \eta + 1)}{2\mathcal{G}_\tau \eta (Denom - Num)} = \frac{\mathcal{G}_\tau + \eta}{2\mathcal{G}_\tau \eta} - \frac{Num(\eta - 1)(\mathcal{G}_\tau - 1)}{2\mathcal{G}_\tau \eta (Denom - Num)}$$

Hence, the following condition is more than sufficient for $\frac{\partial \frac{\mu_L}{\mu_\tau}}{\partial M} > 0$ to hold:

$$\frac{y_L^\eta}{\sum_i y_{it}^\eta} < \frac{\mathcal{G}_\tau + \eta}{2\mathcal{G}_\tau \eta} + \frac{Num(1 - \eta)(\mathcal{G}_\tau - 1)}{2\mathcal{G}_\tau \eta (Denom - Num)}$$

If we only consider cases with $M > 2$, then $\frac{y_L^\eta}{\sum_i y_{it}^\eta} < \frac{1}{2}$, and a sufficient condition is:

$$1 < \frac{\mathcal{G}_\tau + \eta}{\mathcal{G}_\tau \eta} + \frac{Num(1 - \eta)(\mathcal{G}_\tau - 1)}{\mathcal{G}_\tau \eta (Denom - Num)}$$

This holds for any value $0 \leq \eta \leq 1$, since it holds for $\eta = 0, 1$ and for $\eta > 0$, the RHS is monotonically declining because $\frac{\partial \frac{\mathcal{G}_\tau + \eta}{\mathcal{G}_\tau \eta}}{\partial \eta} = \frac{\mathcal{G}_\tau \eta - (\mathcal{G}_\tau + \eta)\mathcal{G}}{(\mathcal{G}_\tau \eta)^2} = -\frac{1}{\eta^2}$. Hence, we always have that $\frac{\partial \frac{\mu_L}{\mu_\tau}}{\partial M} > 0$.

C Assumptions on the productivity volatility process

The definition of the steady state implies that aggregate variables are stable, despite a stochastic process on firm-level productivity. This appendix section describes the assumptions I make on the productivity volatility process. I will be referring to the types of firm that are listed in Lemma 1: low productivity firms (which are always unconstrained in steady state), and all types of high productivity firms, both constrained and unconstrained. I will denote the low productivity firms by type L, the unconstrained high-productivity firms by type H^U, and the constrained firms of type τ, where τ measures the number of periods since a firm's most recent positive productivity shock.

A necessary and sufficient condition for the economy to be in steady state, is that the number of firms of each type is constant for all t. In that case, we immediately have that for all t, $F(a(t+1), z(t)) = F(a', z)$ and hence it is clear from the firm decision rules in (52) and the labor market clearing condition (15) that the other aggregate variables $(K_t, \frac{P_t}{w_t})$ and $H(a,k)$ are constant as well. If the number of firms of each type is not constant over time, then necessarily $F(a(t+1), z(t)) = F(a', z)$ is not constant and the economy is not in steady state.

Since the law of large numbers does not hold under a finite M, I will make additional assumptions on the productivity volatility process to ensure that the economy can still be in steady state as defined in (17). Specifically, I will assume that the assignment of productivity shocks is such that, if a certain transition probability Pr_{xy} to go from state x to state y applies to a set of firms of size N_x, then exactly $Pr_{xy}N_x$ firms will transition from state x to state y.[49] What remains to be defined, are the different states x and y.

In the comparative statics exercise in the paper, I am comparing steady states for different values of M. In order to make valid comparisons across different values of M, the productivity volatility process needs to be identical for different M. To then describe a productivity volatility process that is constant across M, it will be useful to keep track of the following implications of Proposition 1. This proposition is demonstrated conditional on the steady state existing for different values of M, as well as the productivity volatility process being identical across M. Hence, if the characteristics of the productivity volatility process are such that the steady state as defined in (17) exists, and that the process is identical across M, then Proposition 1 holds. In order to describe the productivity volatility process, it will be useful to keep in mind the following implications of the model.

- Implication 1: convergence to the optimal level of capital is reached in a finite number of periods and therefore the maximal τ is finite. This is because $\frac{k_H^*}{k_L^*}$ is finite and g_τ does not converge to zero.
- Implication 2: The number of periods it takes for a high productivity firm to become unconstrained is weakly increasing with M.[50] Let therefore T^M denote the number of periods it takes for a high productivity firm to grow out of its financial constraint in a steady state with M firms.
- Implication 3: there are then in total $(T^M + 2)$ types of firms: L, T^M, H^U

The productivity volatility process is then described as follows. Consider a sufficiently high M, $\bar{M}$. Given a specific productivity volatility process, $\bar{M}$ will be the highest value of M considered in the comparative statics on M. Importantly, given implications 1 and 2, we have that $\forall M < \bar{M} : T^M \leq T^{\bar{M}}$.

[49]One could think of the gods setting up a lottery such that exactly $Pr_{xy}N_x$ firms are selected to transition from x to y.

[50]This is because G_τ is weakly decreasing in M and $\frac{k_H^*}{k_L^*}$ is increasing with M.

Based on implication 3, the productivity volatility process will then be defined by transition probabilities across $(T^{\bar{M}}+2)$ "bins" of firms, namely $L, T^{\bar{M}}, H^U_{\bar{M}}$, where $H^U_{\bar{M}}$ denotes the bin with all the unconstrained high productivity firms for $\bar{M}$. Note that this implies that for $M < \bar{M}$, firms might be in e.g. bin $T^{\bar{M}}$ for the definition of their transition probabilities, although they are already unconstrained and thus of type H^U. The transition probabilities across bins are then defined as follows

- Probability to transition from a_L to a_H, i.e. probability to transition from L to $\tau = 1$: P_{LH}. Then, $(1 - P_{LH})$ is probability to remain within L
- Then, for firms with a_H, the transition probabilities are dependent on τ. Conditional on having a_H in period τ, the probability to continue having a_H is $P_{HH\tau}$.
  - Therefore, conditional on having a_H in $\tau = 1$, the probability of still having a_H in $\tau > 1$, is $\prod_{r=1}^{\tau-1} P_{HHr}$
  - Then, the unconditional probability of having a firm in bin $\tau > 1$ is $P_{LH}\prod_{r=1}^{\tau-1} P_{HHr}$
- Finally,the transition probability of moving from bin $H^U_{\bar{M}}$ to bin L, is P_{HL}.

By specificying these bins, and making the transition probabilities between high and low productivity specific to a bin, I have assured that the number of firms in each bin is stable across periods. This can be seen from the following.

- Denote the number of firms in L by M_L
- Number of firms in bin τ: $M_L P_{LH}\prod_{r=1}^{\tau-1} P_{HHr}$
- Number of firms in $H^U_{\bar{M}}$ can be found by setting the number of exiters from $H^U_{\bar{M}}$ equal to the number of entrants in $H^U_{\bar{M}}$: $M_L P_{LH}\prod_{r=1}^{T^{\bar{M}}} P_{HHr} = P_{HL}M_H$. Hence

$$M_H = \frac{M_L}{P_{HL}} P_{LH} \prod_{r=1}^{T^{\bar{M}}} P_{HHr}$$

- One can then also observe that the number of entrants in L equals the number of exiters from L:

$$P_{LH}M_L = P_{LH}M_L\left[(1-P_{HH1}) + \sum_{\tau=2}^{T^{\bar{M}}}(1-P_{HH\tau})\prod_{r=1}^{\tau-1}P_{HHr}\right] + M_H P_{HL}$$

$$P_{LH}M_L = P_{LH}M_L\left[(1-P_{HH1})\sum_{\tau=2}^{T^{\bar{M}}}(1-P_{HH\tau})\prod_{r=1}^{\tau-1}P_{HHr}\right] + \frac{M_L}{P_{HL}}P_{LH}\prod_{r=1}^{T^{\bar{M}}}P_{HHr}P_{HL}$$

$$1 = (1-P_{HH1}) + \sum_{\tau=2}^{T^{\bar{M}}}\left[(1-P_{HH\tau})\prod_{r=1}^{\tau-1}P_{HHr}\right] + \prod_{r=1}^{T^{\bar{M}}}P_{HHr}$$

Now, note first that $\prod_{r=1}^{T^{\bar{M}}} P_{HHr}$ is the probability, conditional on a firm moving from L to H productivity, that after $T^{\bar{M}}$ periods it still has H productivity. Then note that $\sum_{\tau=1}^{T^{\bar{M}}}(1-P_{HH\tau})\prod_{r=1}^{\tau-1} P_{HHr}$ is the probability that a firm moves back to low productivity at some point before $T^{\bar{M}}$. Hence we always have that

$$1-\prod_{r=1}^{T^{\bar{M}}} P_{HHr} = (1-P_{HH1}) + \sum_{\tau=2}^{T^{\bar{M}}}(1-P_{HH\tau})\prod_{r=1}^{\tau-1} P_{HHr}$$

In other words, the condition for stability of the share of number of firms is always satisfied.[51]

While the described productivity volatility process will ensure that the number of firms in each bin is stable over time, it does not necessarily imply that the number of firms in a bin is an integer. Hence, one has to impose additional restrictions on the values of M under consideration, or the transition probabilities. One such possible restriction is to set $\forall\tau P_{HH\tau} = P_{LH} = P_{LH} = \frac{1}{n}$ and $M_L = n^{(T^{\bar{M}}+x)}$ with $x \geq 2$ and $n \in \mathbb{N}$. This implies that the number of firms in any bin τ is $n^{-\tau}M_L = n^{(T^{\bar{M}}+x-\tau)}$ and in bin H it is n^x.

D Model with young firms

Agents The worker side of the model is unaltered from the baseline model. On the firm side, there continues to be an exogenous, finite set M of firm-owners. In this version of the model, heterogeneity across firms arises from the date at which they are born. Before the start of each period, qM new firms are born with capital levels $k_0 \equiv \zeta\frac{K}{M}$, where K is aggregate capital and $0 < \zeta < 1$. At the same time, a set of firms qM dies before the start of the period, such that the total number of firms remains constant.[52]

Firm-owner i has the following intertemporal preferences at time t:

$$U_{it} = \sum_{s=t}^{\infty}(q\beta)^{s-t}d_{is}$$

Where β is the discount factor, q is the ex-ante probability a firm dies in any given period and d_{it} is firm-owner consumption.

Production of varieties Each firm produces a variety i with a Cobb-Douglas production function, using capital k_{it} and labor l_{it} as inputs. There is no variation in productivity across firms.

$$y_{it} = k_{it}^{\alpha}l_{it}^{1-\alpha} \tag{50}$$

Investment $k_{it+1} = x_{it} + (1-\delta)k_{it}$ is modeled exactly as in the baseline model. The same holds for the definition of the final good, firm-level demand (3), the price index (4), the budget constraint (5), and the financial constraint (6).

D.1 Market structure and optimization in steady state

The market structure and firm-problem are equivalent to the set-up in the baseline model, except that there is no firm-level productivity volatility to be taken into account. Since firms play a one-period game of quantity competition, each firm i sets a quantity y_{it+1} for sale, conditional on the quantities chosen by the other firms in the economy. As discussed in the

[51] Note that $\sum_{\tau=2}^{T^{\bar{M}}}(1-P_{HH\tau})\prod_{r=1}^{\tau-1} P_{HHr} = 1-P_{HH1}+(1-P_{HH2})P_{HH1}+(1-P_{HH3})P_{HH1}P_{HH2}+....+(1-P_{HHT^{\bar{M}}})\prod_{r=1}^{T^{\bar{M}}-1} P_{HHr}$, which confirms the equality.

[52] The ex-ante probability that any firm dies is constant at q, but this probability is not independent across firms as I assume that each period the dying firms hold the same fraction of aggregate capital.

previous subsection, firms make decisions about l_{it+1}, k_{it+1} in period t, given the budget constraint $P_t(k_{it+1} + d_{it}) \leq z_{it}$. Therefore, any firm i's optimal decisions are $k_{it+1}\left(z_{it}, \mathbf{y_{-it+1}}\right)$, $l_{it+1}\left(z_{it}, \mathbf{y_{-it+1}}\right)$, where (z_{it}) characterizes the state for firm i and $\mathbf{y_{-it+1}}$ is the vector of decisions on y_{jt+1} for all $j \neq i$. Through the production function (50), the choice of k_{it+1}, l_{it+1} determines y_{it+1} and thereby $p_{it+1}(y_{it+1}, \mathbf{y_{-it+1}})$ as firms incorporate the demand function (3) into their optimization. As such, this setting entails the following intertemporal problem for the firm, where $\pi_{it}(k_{it}, l_{it}, \mathbf{y_{-it}}) \equiv p_{it}(y_{it}, \mathbf{y_{-it}})y_{it} - w_t l_{it}$:

$$\begin{aligned} \max_{d_{it}, k_{it+1}, l_{it+1}} \mathcal{L} = & \sum_{t=s}^{\infty} E_s\left[\beta^{t-s} d_{it}\right] + \\ & \sum_{t=s}^{\infty} E_s\left[\lambda_{it}\left(\pi_{it}(k_{it}, l_{it}, \mathbf{y_{-it}}) + P_t\left[(1-\delta)k_{it} - k_{it+1} - d_{it}\right]\right) + \Phi_{it}(d_{it})\right] \end{aligned} \tag{51}$$

Since each firm's decision on y_{it+1} depends on $(z_{it}, \mathbf{y_{-it+1}})$, $\mathbf{y_{it+1}}$ will be determined by $F(z(t))$, the distribution of z_{it}, and by the conditions in the labor and goods market implied by M, L.

$$\begin{aligned} k_{it+1}\left(z_{it}, F(z(t)), M, L\right) \\ l_{it+1}\left(z_{it}, F(z(t)), M, L\right) \end{aligned} \tag{52}$$

From here on, the optimization is exactly as in the baseline model, with equivalent expressions for the demand elasticity, the labor choice and the capital choice.

D.2 Steady state equilibrium

An *equilibrium* consists of a set of prices P_t, w_t, p_{it}, a set of consumption $d_{it}(z_{it}, F(z(t)))$, capital $k_{it+1}(z_{it}, F(z(t)))$ and labor $l_{it}(z_{it-1}, F(z(t-1)))$ decisions by firm-owners and consumption by workers $\frac{w_t}{P_t}L$ that satisfy

- the labor market clearing condition

$$L = \sum_{i=1}^{M} l_{it} \tag{53}$$

- the goods market clearing condition

$$Q_t = \sum_{i=1}^{M}(x_{it} + d_{it}) + \int_{l \in L} c_{lt} dl \tag{54}$$

- the optimality conditions for labor and capital for each firm i, conditional on the choices of l_{jt}, k_{jt} of all firms $j \neq i$.
- market-clearing for each variety i: $y_{it} = q_{it}$, satisfying the expression for firm demand.
- the equalized budget constraint $P_t(k_{it+1} + d_{it}) = z_{it}$, and the financial constraint $d_{it} \geq 0$.
- Firms are born with a capital level k_0. This capital level k_0, with $k_0 = \zeta k^*$, where k_0 is inherited from the dead firms, such that necessarily :$\bar{k} \geq k_0$. And here, $\bar{k} = \frac{K}{M}$. In steady state, we know that $k_0 < k_1$ (i.e. since K is constant, all firms are born with the same k_0 and afterwards grow their capital.

D.3 Steady state conditions

- $K_t = K$
- $P_t/w_t = P/w$
- $F(z(t)) = F(z)$,

An implication of $K_t = K$ is that capital growth by surviving firms will have to equal the capital loss from firms dying.

D.3.1 Labor and capital decisions in steady state

It will again be convenient to characterize the solution to the firm's optimization problem by taking the perspective of the cost-minimization problem given the optimal markup characterized in (9). The cost-minimization problem implies the following optimal labor demand in steady state:

$$l_{it} = \left(\frac{(1-\alpha)}{\mu_{it}} \frac{P}{w} \left(\frac{Q}{M} \right)^{1-\eta} k_{it}^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} \tag{55}$$

There are two cases for the firm's capital choice: either $\Phi_{it} = 0$, or $\Phi_{it} > 0$.

Unconstrained firms First consider the case where a firm has $\Phi_{it} = 0$.

$$k^* = \mu_U^{\frac{1}{\eta-1}} \frac{Q}{M} \left(\frac{P(1-\alpha)}{w} \right)^{\frac{\eta-\alpha\eta}{1-\eta}} \left(\frac{\alpha}{r_{it}} \right)^{\frac{1+\alpha\eta-\eta}{1-\eta}} \tag{56}$$

with the new definition $r_{it} \equiv \left(\frac{1}{q\beta} + \delta - 1 \right)$ and μ_U the markup of the unconstrained firm.

Constrained firms When the financial constraint binds, i.e. $\Phi_{it} > 0$. Capital grows as allowed by the budget constraint

$$k_{it+1} = (1-\delta)k_{it} + \left(\left(\frac{(1-\alpha)}{\mu_{it}} \right)^{\frac{\eta-\alpha\eta}{1+\alpha\eta-\eta}} - \left(\frac{(1-\alpha)}{\mu_{it}} \right)^{\frac{1}{1+\alpha\eta-\eta}} \right) \left(\frac{P}{w} \left(\frac{Q}{M} \right)^{1-\eta} k_{it}^{\alpha\eta} \right)^{\frac{1}{1+\alpha\eta-\eta}} \tag{57}$$

This will then imply the following lemma for the capital distribution $H(k)$ in steady state, where τ is the number of periods since the firm was born:

Lemma 3. *Steady state H(k) is given by:*

- *When $\Phi_\tau = 0$, then $k_{i\tau} = k^*$*
- *When $\Phi_\tau > 0$, then $k_{i\tau} = G_\tau k_0$, where $G_\tau = \Pi_{s=0}^{\tau-1}(1+g_s)$ and $g_s = \frac{k_{is+1}}{k_{is}}$*

This way, the capital distribution in this economy is essentially isomorphic to the distribution of the baseline model. Furthermore, the other elements of the system of equations - the markup distribution, TFP, K, $\frac{P}{w}$, Ω - are isomorphic as well, after properly adjusting for the constant productivity. Therefore, this model exhibits analogous comparative statics on M as the baseline model.

E Markup Measurement

The markup measurement is based on De Loecker and Warzynski (2012), who elaborated on the framework introduced by Hall (1986). The main structural assumption for this markup measurement is cost-minimization by firms. Therefore, setup the Lagrangian for cost-minimization on the variable inputs $X_{it}^1, ..., X_{it}^V$,:

$$L_{it}(X_{it}^1, ..., X_{it}^V, K_{it}) = \sum_{v=1}^{V} P^{X_{it}^v} X_{it}^v + r_{it}K_{it} + \lambda_{it}(Q_{it} - Q_{it}(X_{it}^1, ..., X_{it}^V, K_{it}))$$

$$\text{FOC}: \frac{\partial L_{it}}{\partial X_{it}^v} = P^{X_{it}^v} - \lambda_{it}\frac{\partial Q_{it}(.)}{\partial X_{it}} = 0 \Rightarrow \frac{P_{it}^Y}{\lambda_{it}} = \frac{\partial Q_{it}(.) X_{it}^v}{\partial X_{it} Y_{it}} \frac{P_{it}^Y Y_{it}}{P^{X_{it}^v} X_{it}^v}$$

Which implies:

$$\mu_{it} = \frac{\theta_{it}^{X^v}}{\alpha_{it}^X}$$

- Markup $\mu_{it} \equiv \frac{P_{it}^Y}{\lambda_{it}}$,
- the output elasticity for X^v: $\theta_{it}^{X^v} \equiv \frac{\partial Q_{it}(.)}{\partial X_{it}} \frac{X_{it}}{Q_{it}}$
- X's expenditure share in total revenue $\alpha_{it}^{X^v} \equiv \frac{P^{X_{it}^v} X_{it}^v}{P_{it} Q_{it}}$.
  - Note that $\mu_{it} = \frac{\theta_{it}^{X^v}}{\alpha_{it}^X}$ holds for any variable input X_{it}.
  - In the majority of the empirical estimations, I use labor as the variable input. In that case, I define $\alpha_{it}^L \equiv \frac{VA_{it}}{w_t l_{it}}$, where VA_{it} is value added.
  - In some robustness checks, I employ materials as the variable input. In that case, I define $\alpha_{it}^M \equiv \frac{S_{it}}{p_t^M M_{it}}$, where S_{it} is sales and $p_t^M M_{it}$ is expenditure of materials.
- For Cobb-Douglas, θ_{it}^X is constant, so all within-sector variation is driven by α_{it}^X.

F Further stylized Facts

F.1 Robustness on MRPK dispersion and productivity volatility

In this section, I follow Asker et al. (2014) and implement their plant-level robustness check for examining the relationship between MRPK dispersion and productivity volatility. In general, the relationship here is in line with the findings in Asker et al. (2014).

Table A.1: MRPK dispersion: plant-level robustness

	$MRPK_{irst}$ (Gross Revenue)								$MRPK_{irst}$ (Value Added)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
$a_{irst} - a_{irst-1}$	0.723**		0.666**		0.565**		0.441**		1.091**		1.053**		0.997**		0.935**	
	(0.0182)		(0.0149)		(0.00490)		(0.00408)		(0.00613)		(0.00300)		(0.00325)		(0.00332)	
$a_{irst-1} - a_{irst-2}$		0.527**		0.469**		0.239**		0.172**		0.612**		0.564**		0.250**		0.187**
		(0.0174)		(0.0163)		(0.00481)		(0.00480)		(0.0158)		(0.0153)		(0.00592)		(0.00601)
a_{irst-1}	0.906**		0.832**		0.660**		0.515**		1.136**		1.087**		1.017**		0.944**	
	(0.0156)		(0.0117)		(0.00533)		(0.00471)		(0.00826)		(0.00423)		(0.00380)		(0.00387)	
a_{irst-2}		0.756**		0.682**		0.309**		0.230**		0.863**		0.804**		0.325**		0.252**
		(0.0151)		(0.0139)		(0.00624)		(0.00603)		(0.0140)		(0.0139)		(0.00756)		(0.00764)
k_{irst}			-0.199**				-0.479**				-0.134**				-0.241**	
			(0.00491)				(0.00480)				(0.00621)				(0.00296)	
k_{irst-1}				-0.187**				-0.254**				-0.141**				-0.226**
				(0.00560)				(0.00843)				(0.00773)				(0.00906)
Plant FE	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Observations	235765	161861	235765	161861	235765	161861	235765	161861	235765	147713	235765	147713	235765	147713	235765	147713

Standard errors in parentheses

SEs clustered at the sector-level for specifications 1-4, 9-12 and at the plant-level for specifications 5-8, 13-16.

* $p < 0.05$, ** $p < 0.01$

F.2 Capital growth for young plants

Table A.2: Capital growth as a Function of Age

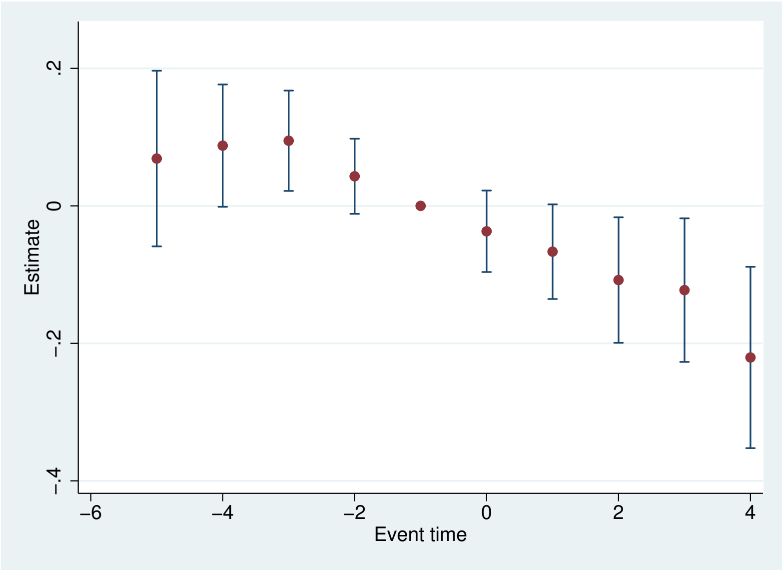
	Plant-level Capital Growth $g(k_{irst})$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\frac{1}{age_{irst}}$	0.0814**	0.140**						
	(0.00412)	(0.00569)						
$\ln(\frac{1}{age_{irst}})$			0.0136**	0.0357**				
			(0.00108)	(0.00169)				
$1(age_{irst} \leq 5)$					0.0597**	0.0576**		
					(0.00210)	(0.00281)		
$1(age_{irst} < 10)$							0.0369**	0.0306**
							(0.00190)	(0.00259)
State-sector-year FE	Yes	No	Yes	No	Yes	No	Yes	No
Plant FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	644922	644922	644922	644922	658886	658886	658886	658886

Standard errors, clustered at the plant level, in parentheses (* $p < 0.05$, ** $p < 0.01$).
Indices are i for plant, r for state, s for sector and t for year.
$g(k_{irst}) = \ln K_{irst+1} - \ln K_{irst}$, where capital is the book value of assets, measured at the start (t) and end $(t+1)$ of the year.

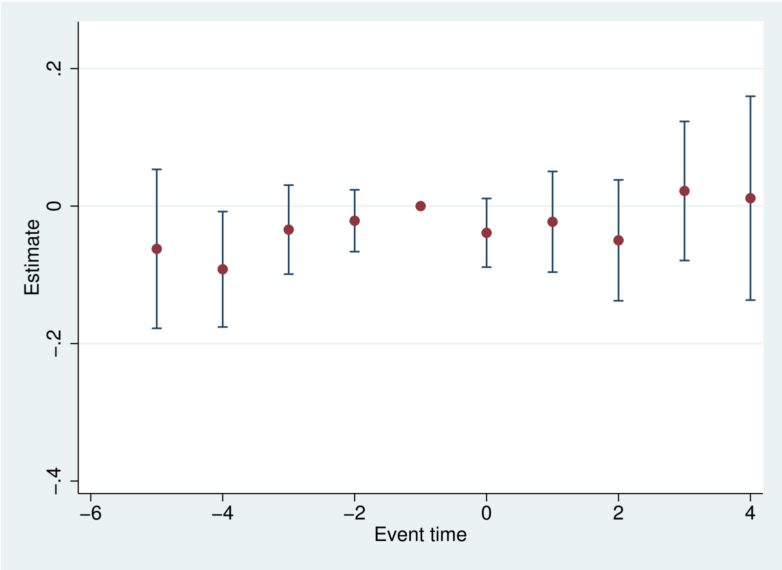
G The effect of dereservation on markup dispersion

Figure A.1: Event-study for the effect of dereservation on markups

(a) Plants with above-median initial markup



(b) Plants 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: $\ln \mu_{irt} = \alpha_i + \gamma_{rt} + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{irt} + \tau] + \varepsilon_{irt}$. Here, α_i is a plant fixed-effect and γ_{rt} is a state-year fixed-effect. I define the time at which the main product of plant i is dereserved as e_{irt}. I impose the normalization that $\beta_{-1} = 0$, and cluster standard errors at the plant-level. Panel (a) shows results for the subset of plants with initial markups weakly above the median markup, and Panel (b) for the other plants. The initial markup is averaged over event times $\tau = -5$ till $\tau = -3$, and the median initial markup is computed after controlling for sector and year fixed effects.