PEDL Research Papers

This research was partly or entirely supported by funding from the research initiative Private Enterprise Development in Low-Income Countries (PEDL), a Department for International Development funded programme run by the Centre for Economic Policy Research (CEPR).

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

Simon Galle *

BI Norwegian Business School

October 2018

Abstract

This paper demonstrates a dual impact of increased competition on misallocation in a setting with both oligopolistic competition and financial constraints. Without financial constraints, more competition unambiguously increases aggregate output by reducing markup levels and markup dispersion. However, with financial constraints, increased competition reduces the profitability of constrained firms, and thereby slows down their rate of self-financed investment and convergence to their optimal capital levels. I test these theoretical predictions by leveraging the pro-competitive impact of India's 1997 dereservation reform. As predicted, this reform leads to reduced markup levels and markup dispersion, and to slower capital convergence.

*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, Daniel Haanwinckel, Ben Handel, Alfonso Irarrázabal, Yusuf Mercan, Plamen Nenov, Louis Raes, Marco Schwarz, Yury Yatsynovich and Moises Yi for many helpful discussions and suggestions. I am deeply thankful to Ishani Tewari for sharing data on India's dereservation reform. 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, Gothenburg, HEC Lausanne, HEC Paris, IHEID Geneva, Johns Hopkins SAIS, KULeuven, NHH Bergen, PEDL-CEPR, Surrey, Tilburg, UC Berkeley, UCLA Anderson and University of Oslo. All errors are my own.

1 Introduction

Aggregate productivity is central to understanding why some countries are rich while others are poor. Since plant-level marginal productivities tend to be much more misaligned in poorer countries, resource misallocation has become a prominent candidate for explaining differences in countries' aggregate productivity.[1] While the potential factors contributing to misallocation are varied, the predominant view in the literature is that competition would be a beneficial force in reducing misallocation. After all, it is highly intuitive that competition will help shift resources from low-performing to high-performing plants, for instance by reducing markup dispersion (Peters, 2016), or by enhancing selection of high-productivity firms.

While the mechanisms driving competition's beneficial impact on aggregate productivity are undeniable, these positive mechanisms do not seem to cover the full story. Consider India, an economy where the original Hsieh and Klenow (2009) analysis found misallocation to be strongly persistent over time. Strikingly, Bollard, Klenow, and Sharma (2013) document a null-effect for most of that country's liberalization reforms, including an extensive licensing reform and a trade liberalization, on the degree of allocative efficiency in its manufacturing sector.[2] This raises the question why the intensified competition associated with these reforms did not lead to a reduction in misallocation.

This paper explores whether the presence of financial constraints can explain this finding. As is well known, limited access to finance is pervasive across developing countries (Levine, 2005), and in India even large firms tend to be credit constrained (Banerjee and Duflo, 2014). In such a context, many firms rely on retained earnings to finance their investment, such that profit levels determine how fast firms are able to save themselves out of their financially constrained position. Hence, by undermining firms' capacity to self-finance their investment in this setting, pro-competitive reforms may not have the expected effect of reducing misallocation.

Theoretically, I formalize this argument by combining two benchmark models of misallocation, namely the model of "capital misallocation" from Midrigan and Xu (2014) and the model for "markup misallocation" from Atkeson and Burstein (2008). In the absence of financial constraints, intensified competition reduces misallocation in an oligopolistic sector, by lowering markups toward their lower bound and thereby depressing markup dispersion. While this beneficial impact of competition on markup misallocation remains central in my framework, I demonstrate that the introduction of financial constraints crucially leads to a second, harmful impact of competition on misallocation.

In the model, firms experience random shocks to their idiosyncratic productivity, and after a positive productivity shock, they optimally choose to grow their capital stock. Critically though, their limited access to finance hampers their ability to do so. Since financially constrained firms then rely on retained earnings to finance their investment, their rate of self-financed capital growth becomes a function of their optimal markup. Increased competition, by reducing firms' markups, negatively affects their speed of capital convergence in response to a positive productivity shock. This way, competition amplifies the difference between a constrained firm's

[1] For the case of India, which will be the country of interest for this paper, Bils, Klenow, and Ruane (2017) argue that misallocation of resources, after controlling for potential measurement error, accounts for 30 to 40% of the difference in aggregate manufacturing output per capita between the United States and India.

[2] The one exception is the financial liberalization reform, which improved allocative efficiency. Critically though, in contrast to the other reforms, this type of reform affects competition in the manufacturing sector at most indirectly. At the same time, it appears to have reduced preferential treatment of certain firms (Bhaduri, 2005), which directly affects misallocation.

optimal and actual level of capital, i.e. the "capital wedge," and worsens capital misallocation. Interestingly, I derive these analytical results on the dual impact of competition, namely that it reduces markup misallocation but amplifies capital misallocation, in a setting where there is no closed-form solution for the distribution of markups and capital.

In the second part of the paper, I test and confirm the predictions of the model in the context of the Indian manufacturing sector. To this end, I use a natural experiment arising from the staggered implementation of an industrial policy reform, namely the 1997 dereservation episode. This reform removed the investment ceilings imposed for the production of certain product categories, which led to the entry of new, larger firms in the production of the now dereserved product categories. Hence, the reform exposed incumbent plants to stiffer competition. Empirically, I examine the impact of the reform on incumbents' markups and their capital convergence. First I demonstrate in an event study that the dereservation reform led to lower markups for incumbent plants, which confirms the pro-competitive impact of the reform. Moreover, markups for plants with an initially higher markup fell more than for plants with a lower initial markup. Hence, the reform reduced markup dispersion, in line with the theory's predictions.

Next, I turn to testing the novel prediction of my model, namely the negative impact of competition on capital convergence. Since a plant's optimal level of capital is unobserved, I focus on convergence in marginal revenue product of capital (MRPK), inspired by Asker, Collard-Wexler, and De Loecker (2014). I proxy for a plant's optimal MRPK with a flexible function, including a plant fixed-effect to allow for maximal cross-plant heterogeneity, and find that plants exhibit strong and robust convergence to this measure of optimal MRPK. This enables me to use the speed at which a plant converges back to its optimal MRPK as an empirical counterpart for the model's speed of convergence to optimal capital levels. I then find that MRPK convergence is indeed slower after a plant's products have been dereserved.

To strengthen the external validity of the empirical analysis, I also study MRPK convergence on the full panel of plants, since I can only examine the pro-competitive impact of the dereservation reform on the subset of incumbent plants described above. For the full panel, the measure of competition will be the median markup across plants observed in the same state, sector and year, and this median value is plausibly exogenous from the perspective of the individual plant. I document that a higher median markup, indicating less competition, is associated with faster plant-level MRPK convergence. To further corroborate the theoretical mechanism, I then explore heterogeneity in the impact of competition along the degree of plants' financial dependence. Using the standard Rajan and Zingales (1998) measures, I find that plants in sectors with higher degrees of financial dependence exhibit a stronger sensitivity to the degree of competition.

The measurement in this first battery of empirical tests relies on standard Cobb-Douglas assumptions for the MRPK values, and on an autoregression framework for the speed of convergence. The advantage of this empirical approach is that it links closely to the model with productivity volatility, but a disadvantage is that this framework may be less transparent than a more reduced-form approach. For this reason, I implement all the above tests also on capital growth for young plants, the benefit being that capital growth is a reduced-form object in the data. This empirical strategy is motivated by a version of the model where newborn firms are undercapitalized, and therefore competition will slow down their capital growth. Here, the assumption that newborn firms are undercapitalized is in line with the stylized facts in the data. For all three tests, namely the impact of dereservation on incumbent plants, the impact of the median markup on the full panel of plants, and for the heterogeneity along financial dependence, the evidence once more lines up with the model's predictions.

This paper is part of the burgeoning literature on resource misallocation, started by the seminal contribution of Restuccia and Rogerson (2008), and reviewed by Restuccia and Rogerson (2013) and Hopenhayn (2014). In addition to the work mentioned earlier, my paper is closely related to Moll (2014) and Itskhoki and Moll (2018), who also analyze capital misallocation analytically.[3] However, they do so in a setting of perfect competition, whereas the key contribution of this paper is to study capital misallocation under oligopolistic competition. Interestingly, the oligopolistic competition setting implies that the closed-form results from Moll (2014) are no longer applicable here. Where a standard approach would then rely on simulations, I am still able to derive analytical results.

Modeling oligopolistic competition as in Atkeson and Burstein (2008) is an increasingly common theoretical strategy to examine variations in competition at the macroeconomic level, see for instance Edmond, Midrigan, and Xu (2015); Brooks, Kaboski, and Li (2016) and Hottman, Redding, and Weinstein (2016).[4] None of these papers feature financial frictions.

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

Taken together, the contribution of this paper is to develop a more nuanced understanding of the positive, as well as the underexamined negative impact of competition on misallocation.[5] The message of this paper is not that the benefits of competition should be discarded and liberalization should be abandoned. Instead, this paper suggests that in order to maximize the gains from liberalization, it is critical to first optimize access to credit throughout the economy.

[3]In a recent contribution, Jungherr and Strauss (2017) confirm the role of market power in facilitating growth for the Korean manufacturing sector, while Ventura and Voth (2015) show how sovereign debt may have accelerated the industrial revolution by depressing factor prices and increasing entrepreneurs' profitability. Next, in a model of technology choice, Foellmi and Oechslin (2016) study the negative impact of increased competition due to international trade on credit access. Buera, Kaboski, and Shin (2015) provide a broad overview of the role of financial constraints in macro development. Then, in a business cycle context, Buera and Nicolini (2017) and Caggese and Pérez-Orive (2017) model the contribution of capital misallocation to liquidity traps and secular stagnation, respectively. Finally, Gopinath, Kalemli-Özcan, Karabarbounis, and Villegas-Sanchez (2017) document how increased capital misallocation in Southern Europe was associated with a decline in the real interest rate, and Kehrig and Vincent (2017) examine how within-firm credit allocation affects the cross-plant dispersion in MRPK.

[4]Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2017), Peters (2016) and Mrázová, Neary, and Parenti (2017) provide alternative frameworks for analyzing the markup distribution, while De Loecker and Eeckhout (2017) focus on the *level* of markups in their analysis.

[5]In the literature on industrial organization, it is well-established that increasing competition can have both positive and negative effects on aggregate output or welfare (see e.g. Mankiw and Whinston (1986); Aghion, Bloom, Blundell, Griffith, and Howitt (2005); Gilbert (2006); Aghion, Akcigit, and Howitt (2014)). 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 underexamined.

I now proceed by presenting the theory. Then, Section 3 discusses the data and Section 4 examines the impact of the industrial policy reform. Section 5 analyzes the impact of competition for the full panel, Section 6 studies capital convergence for young plants and Section 7 concludes.

2 Theory

2.1 Setup of the economy

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

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

where S is the set of sectors and Q_{st} is a sector-level composite good for sector s in period t. The standard price index P_t^F for the final good is $\ln P_t^F = \int_{s\in S} \phi_s \ln(P_{st}/\phi_s)ds$, where P_{st} is the price index for sector s. A direct implication of this setup is that the optimal expenditure shares on goods for sector s are constant at $\phi_s = \frac{P_{st}Q_{st}}{P_t^F Q_t^F}$.

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

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

where q_{ist} is consumption of the variety from firm i in sector s at time t, σ is the elasticity of substitution, with $\sigma > 1$, and M_s is the number of firms in sector s. The fact that a sector's CES aggregate consists of a finite number of firms, as in Atkeson and Burstein (2008), implies that the intensity of competition is a function of that number of firms. The term $M_s^{\frac{1}{1-\sigma}}$ eliminates love of variety, as in Blanchard and Kiyotaki (1987) or Jaimovich (2007). In the analysis below, this elimination of love of variety allows me to isolate the pro-competitive effects of changes in M_s. The inverse demand function and associated revenue function v_{ist} for variety i are then given by:

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

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

where the sectoral price index is

$$P_{st} = M_s^{\frac{1}{\sigma-1}} \left(\sum_{i=1}^{M_s} p_{ist}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}. \tag{4}$$

In the oligopolistic setting under consideration, a firm's demand will become more inelastic as its market share m_{ist} increases:[6]

[6]While the precise value of the demand elasticity will depend on the details of the oligopolistic market structure, the qualitative relation between market share and demand elasticity in equation (5) holds under both Bertrand and Cournot competition. In case firms engage in Cournot competition, their demand elasticity is $\varepsilon_{ist}(q_{ist}) =$

$$\varepsilon_{ist}(m_{ist}) \equiv -\frac{\partial q_{ist}}{\partial p_{ist}}\frac{p_{ist}}{q_{ist}} \quad \text{with} \quad \frac{\partial \varepsilon_{ist}(m_{ist})}{\partial m_{ist}} < 0, \quad \text{and} \quad 1 \leq \varepsilon_{ist}(m_{ist}) < \sigma, \tag{5}$$

where the market share is defined as

$$m_{ist} \equiv \frac{v_{ist}}{\sum_{j=1}^{M_s} v_{jst}} = \frac{q_{ist}^{\frac{\sigma-1}{\sigma}}}{\sum_{j=1}^{M_s} q_{jst}^{\frac{\sigma-1}{\sigma}}}. \tag{6}$$

Below we will see how variation in the demand elasticity, given productivity differences between firms, leads to variation in markups across firms, and thereby to "markup misallocation." I derive this result for the benchmark case of Cournot competition, but the analytical results hold for Bertrand competition as well.

The economy has two types of infinitely lived agents: workers and firm-owners. A measure L of workers supplies labor inelastically, and each worker is hired at a wage w_t. Next, in each sector there is an exogenous number M_s of firm-owners. Both workers and firm-owners consume the final good, and have the following intertemporal preferences over their consumption:

$$U_{jt} = \sum_{t=r}^{\infty} \beta^{t-r} c_{jt}, \tag{7}$$

where β is the discount factor and j denotes either a specific worker l, or a certain firm-owner i in sector s. As I explain below, the assumption of linear preferences will ensure that when firm owners are financially constrained, they optimally choose to set their consumption to zero and use all retained earnings for capital growth. This corner solution for consumption allows me to handle the comparative statics on the distribution of firms' capital growth analytically.

Each firm produces y_{ist}, the output for its variety, using capital k_{ist} and labor l_{ist} according to a Cobb-Douglas production function:[7]

$$y_{ist}(l_{ist}, k_{ist}) = z_{ist} k_{ist}^{\alpha} l_{ist}^{1-\alpha}, \tag{8}$$

where each firm's productivity z_{ist} follows a Markov process over the state space $\{z_{sL}, z_{sH}\}$, with $z_{sL} < z_{sH}$, and switching probabilities between the two states are strictly positive.[8] The assumptions on this Markov process are such that, even though there are a finite number of firms in each sector and therefore the law of large numbers does not hold, the number of firms of a given "type" is constant over time (see Appendix B). Here, the definition of a firm's type will become clear after Lemma 2. Together with the financial frictions that I introduce below, the volatility in productivity can lead to some firms being financially constrained, even in steady state, as in Midrigan and Xu (2014) or Moll (2014).

A firm accumulates capital according to the standard equation of motion:

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

where δ is the depreciation rate and where investment x_{ist} is financed at the end of period t.

$\left[\frac{1}{\sigma}(1-m_{ist}) + m_{ist}\right]^{-1}$, and in case of Bertrand competition, it is $\varepsilon_{ist}(q_{ist}) = \sigma(1-m_{ist}) + m_{ist}$ (see Atkeson and Burstein (2008), and Amiti, Itskhoki, and Konings (2016) for derivations.)

[7] In this production function, α is allowed to be sector specific, but I drop the subscript s to ease the notational burden.

[8] Increasing the dimensionality of the state space would add complexity to the analytical solution of the comparative statics, without yielding additional economic insight.

Investment can be funded using retained earnings or debt, since firms can borrow from workers using a one-period risk-free security at an interest rate r_t^d. I assume that a firm's debt d_{ist} is constrained to be weakly positive ($d_{ist} \geq 0$). If instead the firm-owner would be able to save, given her linear preferences, the firm-owner could choose to accumulate sufficient savings such as never to be financially constrained in steady state. On the debt market, firms borrow from workers, who thereby accumulate assets b_{lt}. The equilibrium in the debt market holds when

$$\int_{l \in L} b_{lt} dl = \int_{s \in S} \left(\sum_{i=1}^{M_s} d_{ist} \right) ds.$$

Importantly, the firm's borrowing is subject to a collateral constraint as in Moll (2014), which puts a limit on the firm's leverage ratio:

$$\frac{d_{ist}}{k_{ist}} \leq \lambda, \qquad \lambda \geq 0. \tag{9}$$

In contrast to investment, payments to labor l_{ist} are only made after revenue in period t is realized. Before the end of period t, i.e. after revenue is realized, debt is repaid and capital has depreciated, but before decisions about borrowing, consumption, investment and labor hiring are made, a firm owner's net real wealth is then

$$a_{ist}(l_{ist}, k_{ist}, \mathbf{y_{-ist}}) \equiv \frac{\pi_{ist}(l_{ist}, k_{ist}, \mathbf{y_{-ist}})}{P_t^F} + (1-\delta)k_{ist} - (1 + r_t^d)d_{ist},$$

where $\pi_{ist}(l_{ist}, k_{ist}, \mathbf{y_{-ist}}) \equiv v_{ist}(y_{ist}, P_{st}(y_{ist}, \mathbf{y_{-ist}})) - w_t l_{ist}$, i.e. revenue net of payments to labor, y_{ist} is a function of l_{ist} and k_{ist}, and $\mathbf{y_{-ist}}$ is the vector of output choices of firm i's competitors. Therefore, the owner faces the following period-by-period budget constraint:[9]

$$k_{ist+1} + c_{ist} \leq a_{ist}(l_{ist}, k_{ist}, \mathbf{y_{-ist}}) + d_{ist+1}, \tag{10}$$

Since there is a non-negativity constraint on firm-owner consumption ($c_{ist} \geq 0$), given the collateral constraint, next period's capital level is bounded above as follows:

$$k_{ist+1} \leq \frac{a_{ist}(l_{ist}, k_{ist}, \mathbf{y_{-ist}})}{1-\lambda}.$$

Hence, for $\lambda \geq 1$, firms are not constrained in terms of their capital choice.

In terms of timing of the decision process, firms' productivities for $t+1$ are revealed at the end of period t, and at that point firms make decisions, subject to their budget constraint, about consumption in period t and labor, capital and debt for period $t+1$. Given this set-up, the individual firm's relevant state variables are future productivity z_{ist+1} and wealth a_{ist}. The state of a firm's competitors can be summarized by $D_{-i}^s(z_{jst+1}, a_{jst})$, the joint distribution of productivity and wealth for all firms in sector s excluding firm i. In terms of notation, $D^s(z_{ist+1}, a_{ist})$ denotes the joint distribution of productivity and wealth for *all* firms in industry s.

Workers' problem Workers optimize their intertemporal utility function from equation (7) subject to their period by period budget constraint

[9] In the subsequent analysis, I only consider cases where any firm owner's initial wealth is strictly positive, i.e. $a_{is0} > 0$, such that the optimization implies that any firm's wealth is positive in any period. In combination with the Inada conditions implied by the revenue function, this ensures that firms always set strictly positive levels of capital and labor.

$$c_{lt} + b_{lt+1} \leq \frac{w_t}{P_t^F} + (1 + r_t^d) b_{lt}. \tag{11}$$

Their linear utility implies the following optimal choices for saving and consumption

$$\begin{aligned}
\left(r_{t+1}^d > \frac{1}{\beta} - 1\right) &\implies \left(b_{lt+1}^* > 0, c_{lt+1}^* = 0\right) \\
\left(r_{t+1}^d < \frac{1}{\beta} - 1\right) &\implies \left(b_{lt+1}^* = 0, c_{lt+1}^* > 0\right) \\
\left(r_{t+1}^d = \frac{1}{\beta} - 1\right) &\implies \left(b_{lt+1}^* \geq 0, c_{lt+1}^* \geq 0\right)
\end{aligned} \tag{12}$$

Market structure and the firm's problem The firms play an infinitely repeated Cournot-type game where they decide each period about investment and labor hiring. In this setting, a strategy ψ_{ist} of a firm consists of a set of decision rules for capital, labor and debt, valid for all current and future periods, that are conditional on the firm's own state, the state of its competitors, the state of the macroeconomy, and the history of the game h_{st}. The state of the macroeconomy is summarized in $\mathbf{F_t} \equiv \{w_t, P_t^F, Q_t^F, r_t^d\}$. For notational convenience, I write these decision rules as a function of $D^s(z_{ist+1}, a_{ist})$, which includes both the firm's own state and the state of its competitors.[10] In each sector, the firms then choose the strategy that maximizes their present value of consumption, conditional on ψ_{-ist}, the strategies of the firm's competitors:

$$\max_{\psi_{ist}} E_t[U_{ist}(\psi_{ist}, \psi_{-ist})], \tag{13}$$

subject to the budget constraint from equation (10) and the collateral constraint from equation (9). This optimization will imply that the budget constraint is satisfied with equality, and therefore decisions for capital and debt immediately imply a decision for consumption as well. In addition, decisions on labor and capital for next period imply a decision for output and associated revenue.

Definition. ***An industry equilibrium*** *consists of a set of strategies ψ_{ist} for all firms i in sector s that constitute a Nash equilibrium given the optimization problem in (13), conditional on a specific path for the macroeconomy $\{\mathbf{F_t}\}$.*

In the next sections, I will discuss the firms' optimization process and the resulting industry equilibrium in greater detail for extreme cases of the collateral constraint λ. For now, I note that the firms' equilibrium decision rules for capital, labor and debt are a function of the state of the sector, the history of the game, and the state of the macroeconomy:

$$\begin{aligned}
&k_{ist+1}^*(D^s(z_{ist+1}, a_{ist}), h_{st}, \mathbf{F_{t+1}}) \\
&l_{ist+1}^*(D^s(z_{ist+1}, a_{ist}), h_{st}, \mathbf{F_{t+1}}) \\
&d_{ist+1}^*(D^s(z_{ist+1}, a_{ist}), h_{st}, \mathbf{F_{t+1}})
\end{aligned} \tag{14}$$

Within each sector, the decision rules on labor and capital, given the state of the sector, result in

[10] While a firm's decision depends on the macroeconomic prices $w_{t+1}, P_{t+1}^F, r_{t+1}^d$, each sector is atomistic, and therefore these factor prices are exogenous to the individual sector. For convenience I will therefore, omit these prices from the notation of the decision rules, except in the definition of general equilibrium below, where the aggregate factor prices play a central role.

the joint distribution of productivity, capital and labor, denoted by $H^s(a_{ist}, k_{ist}, l_{ist})$.

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

$$D^s(z_{ist+1}, a_{ist}) = D^s(z', a),$$

and so is the joint distribution of productivity, capital and labor:

$$H^s(z_{ist}, k_{ist}, l_{ist}) = H^s(z, k, l).$$

Second, workers' decision rules for saving and consumption described in (12) *that satisfy the budget constraint in equation* (11) *with equality. Third, a stable macroeconomic state* $\mathbf{F}$*: a wage* w*, a final good price* P^F *an interest rate* r^d*, such that the labor market clears in every period:*

$$L = \int_{s\in S} \sum_{i=1}^{M_s} l^*_{ist+1}(D^s(z', a), h_{st}, \mathbf{F}) ds, \tag{15}$$

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

$$\int_{l\in L} b^*_{lt+1}(\mathbf{F}) dl = \int_{s\in S} \sum_{i=1}^{M_s} d^*_{ist+1}(D^s(z', a), h_{st}, \mathbf{F}) ds, \tag{16}$$

The interest rate on debt in a steady state equilibrium will be $r^d = \frac{1}{\beta} - 1$. If the interest rate would deviate from this level, then in each period, workers would either only accumulate savings and have zero consumption, or aim to consume only in the present period, both of which cannot be a steady state equilibrium. Also note that since all firms and all workers satisfy their budget constraints with equality, when the labor and debt market for period $t + 1$ clear, then the goods market for period t also clears given the real wage $\frac{w}{P^F}$ determined in the previous period:

$$Q^F - \frac{w}{P^F} L = \int_{s\in S} \sum_{i=1}^{M_s} \left(x^*_{ist}(D^s(a', z), h_{st}, \mathbf{F}) + d^*_{ist}(D^s(a', z), h_{st}, \mathbf{F}) \right) ds.$$

In the next section, I examine two types of steady state equilibria. First I look into the benchmark case where firms can finance any desired level of capital, and afterwards I consider the limit case where firms have no access to external finance.

2.2 No capital constraints

In this section, I set $\lambda = 1$ to examine the benchmark case where firms face no limit on their investment in capital. The collateral constraint then only constrains the firm-owner's consumption level, but imposes no restrictions on her investment in capital. It is convenient to solve the firm's problem via the dual approach of first minimizing costs for any output and then determining optimal output levels. The marginal cost, in terms of utility from consumption in period $t+1$, of the inputs are the real wage w_{t+1}/P_{t+1} for labor, and the rental rate of capital $r^k \equiv \frac{1}{\beta} + \delta - 1$. Given these factor costs, standard cost minimization for Cobb-Douglas production functions implies that firms have the following factor demands:

$$k^*(y_{ist+1}) = \frac{w_{t+1}}{P^F_{t+1}} \frac{1}{r^k} \frac{\alpha}{1-\alpha} y_{ist} \tag{17}$$

$$l^*(y_{ist+1}) = \frac{P^F_{t+1}}{w_{t+1}} r^k \frac{1-\alpha}{\alpha} y_{ist} \tag{18}$$

which result in the following constant marginal cost:

$$MC^u_{st}(z_{ist}) = \frac{1}{z_{ist}} \frac{(r^k)^\alpha \left(\frac{w_{t+1}}{P^F_{t+1}}\right)^{1-\alpha}}{\alpha^\alpha (1-\alpha)^{1-\alpha}}, \tag{19}$$

where the superscript u stands for unconstrained. Given this marginal cost, firms then maximize profits as a function only of output produced, where we can rewrite the profit maximization problem as:

$$\max_{\{y_{ist+1}\}} \sum_{t=v}^{\infty} \beta^{t-v} E_v[v_{ist}(y_{ist}, P_{st}(y_{ist}, \mathbf{y_{-ist}})) - MC^u_{st}(z_{ist}) y_{ist}].$$

This problem is closely related to a standard one-shot Cournot game, except that here the strategic interaction among firms is also dynamic. To understand the potential Nash equilibria, it is useful to first consider the one-period version of this game, where each firm's objective function is to maximize the following profit function: $v_{is}(y_{is}, P_s(y_{is}, \mathbf{y_{-is}})) - MC^u_{st}(z_{ist}) y_{is}$. This optimization implies that each firm sets the optimal markup $\mu_{is}(q_{is}) = \frac{\varepsilon_{is}(q_{is})}{\varepsilon_{is}(q_{is})-1}$. Given the demand function, the firm's best response, or reaction function is then implicitly characterized by:

$$\frac{\varepsilon_{is}(m_{is}(q_{is}))}{\varepsilon_{is}(m_{is}(q_{is})) - 1} = \frac{p_{is}(q_{is})}{MC^u_{st}(z_{ist})}. \tag{20}$$

Collecting the reaction functions for all firms in sector s, we then have M_s equations for M_s unknown values for q_{is}, and the solution to this system of equations is the unique Nash equilibrium for this one-shot Cournot oligopoly game (see also Atkeson and Burstein (2008)).[11]

Returning to the dynamic setting, it is clear that the dynamic game is an infinite repetition of the one-period game, except that firms' productivities are stochastic. Analogous to the argument in Friedman (1971), an infinite repetition of the equilibrium in the one-period game is a subgame perfect equilibrium in the dynamic game. The reason is that when all other firms play the best response from the stage game in every period, a deviation by one firm from this best response in one period lowers the firm's profit in that period, and leaves unaltered its profits in the other periods. In this equilibrium, each firm chooses the strategy where they play the stage game's best response to other firm's output choices in each period, regardless of the history of the game. Since this equilibrium consists of the infinite repetition of the equilibrium strategies from the one-period game, there is necessarily a Nash equilibrium in each subgame of the game. Moreover, since the strategies and actions are indepedent of the history of the game, this subgame perfect equilibrium is also Markov perfect. Of course, as implied by the Folk Theorems, the

[11] Abstracting from factor prices that are exogenous to the individual industry, in this Cournot game a firm's payoff depends on two variables, namely his own output and the output of his competitors as summarizes by the $P_s(\mathbf{y_{ist}})$. It is therefore straightforward to show that this game falls under the class of Aggregative Games, as defined by Corchon (1994) or Acemoglu and Jensen (2013). For this class of games, Corchon (1994) explains the uniqueness of the equilibrium, before examining comparative statics on the number of players.

above equilibrium is not the unique subgame perfect equilibrium, but as a repetition of the stage game equilibrium, it is a natural choice as benchmark equilibrium. I also return to the focus on this specific equilibrium in the next section.[12]

2.3 Markup misallocation

In equilibrium, high-productivity firms have a higher market share than the low-productivity firms.[13] As a consequence, high-productivity firms set higher markups than the low-productivity firms: $\mu_{sH} > \mu_{sL}$, where subscripts H, L henceforth refer to the high- and low-productivity firms' optimal choices. Since this markup dispersion is associated with unequal marginal products across high and low productivity firms, there is within-sector resource misallocation.

A formal way to diagnose this "markup misallocation" is by comparing the optimal to the actual input factor ratios across firms. It is straightforward to show that, for any given M_s, optimal within-sector resource allocation would imply equalized marginal products across firms and therefore the following factor ratios:[14]

$$\frac{\tilde{l}_{sH}}{\tilde{l}_{sL}} = \frac{\tilde{k}_{sH}}{\tilde{k}_{sL}} = \left(\frac{z_{sH}}{z_{sL}}\right)^{\sigma-1},$$

with the tilde referring to the socially optimal input choices. In contrast, given the factor demand functions from equations (17) and (18), firms' best responses in equation (20) result in the following equilibrium factor ratios:

$$\frac{l_{sH}}{l_{sL}} = \frac{k_{sH}}{k_{sL}} = \left(\frac{z_{sH}}{z_{sL}}\right)^{\sigma-1}\left(\frac{\mu_{sL}}{\mu_{sH}}\right)^{\sigma}.$$

Since high-productivity firms set higher markups, it follows that $\frac{\mu_{sL}}{\mu_{sH}} < 1$ as long as M_s is finite. Hence, in the decentralized equilibrium with finite number of firms, the factor ratios are too low compared to the socially optimal allocation. In other words, high-productivity firms are too small relative to the low-productivity firms. However, when competition goes to its upper limit, i.e. when $M_s \to \infty$, firms' market shares become atomistic and each firm's demand elasticity $\varepsilon_{ist}(m_{ist}(q_{ist})) \to \sigma$. Hence, in this limit case we arrive at monopolostic competition, where all firms set an identical, constant markup equal to $\frac{\sigma}{\sigma-1}$. Hence, markup dispersion, and associated markup misallocation, disappears when $M_s \to \infty$.

Lemma 1. *When firms have no capital constraints, taking competition to its upper limit, i.e. $M_s \to \infty$, ensures that marginal products are equalized across firms within sector s.*

This analysis of optimal resource allocation does not take into account the issue of selection. Naturally, comparing across equilibria with only atomistic firms, a sector's productivity is high-

[12]Note that the expression for the reaction functions in equation (20) is robust to the choice of market structure, as they nest the reaction functions from both the Cournot and the Bertrand oligopoly game. The same will hold when I analyze the equilibrium with financial constraints and the related comparative statics.

[13] Consider the ratio of their reaction functions:

$$\frac{y_{sH}}{y_{sL}} = \left(\frac{z_{sH}}{z_{sL}}\frac{\mu_{sL}(m_{sL})}{\mu_{sH}(m_{sH})}\right)^{\sigma}$$

This relation implies that $y_{sH} > y_{sL}$. To see why, suppose to the contrary that $y_{sH} < y_{sL}$. This would require $\mu_{sH} < \mu_{sL}$. Given that $\frac{\partial \varepsilon_{ist}(m_{ist})}{\partial m_{ist}} < 0$, this in turn would imply that $m_{sH} < m_{sL}$, and this results in a contradiction with $y_{sH} < y_{sL}$. Hence, the supposition is false and its opposite must be true.

[14]To obtain this well-known result, one can maximize total industry output Q_{st}, for any given amount of capital and labor available for sector s.

est when only high-productivity firms are producing. The current model is not set up to analyze this type of productivity-enhancing selection.[15] The analysis of markup misallocation here, requires the existence of productivity differences, and such differences typically persist even after the lowest productivity plants disappear through selection. In the next section, financial constraints become binding. In that setting, the key mechanism is that individual firms can become financially constrained due to productivity volatility, which generates within-firm capital growth. Such within-firm growth can also occur in a setting with selection. In fact, Midrigan and Xu (2014) provide a model of financial constraints and productivity volatility, where there is a selection mechanism operating on the firms in the "modern" sector.

2.4 No external finance

In the previous section, we examined the case with no capital constraints. Now, we consider the opposite extreme, where firms have no access to external finance, i.e. $\lambda = 0$.[16] Without access to external financing, each firm's budget constraint simplifies to:

$$k_{ist+1} + c_{ist} \leq a_{ist}(l_{ist}, k_{ist}, \mathbf{y_{-ist}}),$$

where $c_{ist} \geq 0$.

I again start by first performing cost-minimization for any specific output level, denoted by $\bar{y}_{ist}$. For this output level, the firm can either be unconstrained or constrained. Whether a firm is constrained, depends on its maximum capital level, defined as:

$$k^c_{ist+1} \equiv (1-\delta)k_{ist} + \frac{\pi_{ist}}{P^F_t}.$$

If for a given output level $\bar{y}_{ist}$, a firm's capital demand from equation (17) is below k^c_{ist}, then its demand functions for capital and labor are exactly as in the case with no financial constraints, and this gives rise to the standard marginal cost function for a Cobb-Douglas, as in equation (19). In contrast, the firm is constrained for a specific $\bar{y}_{ist}$, when the capital demand from equation (17) is larger than its maximum capital amount. Next, define the return on the constrained capital level, denoted r^{kc}_t, such that the following equality holds:

$$k^c_{ist+1} = \frac{w_{t+1}}{P^F_{t+1}} \frac{1}{r^{kc}_{t+1}} \frac{\alpha}{1-\alpha} \bar{y}_{ist+1},$$

which naturally implies that $r^{kc}_{t+1} > r^k$. In turn, the fact that the actual return on capital is higher than r^k, which is the required return on capital to ensure the intertemporal optimality of investment, entails that it will be optimal for the constrained firm to invest all its wealth in capital by setting it at k^c_{ist+1}. If the firm would invest less, it would have a strictly lower net present value of expected utility. Note that these input demand functions for capital implicitly define the consumption path for firms. When firms are constrained, they invest all their revenue in order to attain k^c_{ist+1}. When they are not constrained, they invest until they reach their optimal capital choice and consume the remaining revenue.

[15] Dhingra and Morrow (2016) provide an extensive analysis of the welfare effects of increased competition and product diversity under firm heterogeneity.

[16] In models with heterogeneous agents and financial constraints, it is becoming increasingly common to analyze the limit case of no external financing to obtain analytical results. See e.g. Krusell, Mukoyama, and Smith Jr (2011); Werning (2015) and Ravn and Sterk (2016) in the context of income risk for consumers. This paper is one of the first to employ this strategy in the context of firm heterogeneity and resource misallocation.

Since the capital level of the constrained firm is at its maximum, at the margin the firm can only adjust its labor input. For any quantity $\bar{y}_{ist}$, its total variable costs are $\frac{w}{P^F} l(\bar{y}_{ist})$, which implies that it has the following marginal cost function:

$$MC^c_{st}(z_{ist}, \bar{y}_{ist}, k^c_{ist}) = \frac{w_t}{(1-\alpha)P^F_t} \left(\frac{\bar{y}^\alpha_{ist}}{z_{ist}(k^c_{ist})^\alpha} \right)^{\frac{1}{1-\alpha}} \tag{21}$$

Combining both the unconstrained and the constrained case, and noting that k^c_{ist} is a function of a_{ist-1}, a firm's marginal cost is a function of its productivity, output level, and wealth in the previous period: $MC_{ist}(z_{ist}, y_{ist}, a_{ist-1})$.

The fact that firms can be financially constrained, implies that their actions depend not only on their productivity, but also on their wealth. This implies that firms may have an incentive to influence the future wealth distribution, thereby affecting other firms' future actions. For instance, firms may expand their production and drive down prices in order to slow down the growth path of its financially constrained competitors. Importantly, when this happens off the equilibrium path, namely constrained firms' capital growth being slowed down, then the joint distribution of capital and productivity becomes non-stationary. This non-stationarity renders the analytical derivation of subgame perfect equilibria highly challenging in the case of financial constraints.

To sidestep this analytical challenge, I focus on the subset of Nash equilibria where firms have the reaction functions from the static game in each period:

Assumption 1. *On the equilibrium path, firms' reaction functions take the following form:*

$$\frac{\varepsilon(m_{ist}(y_{ist}))}{\varepsilon(m_{ist}(y_{ist})) - 1} = \frac{p_{ist}(y_{ist})}{MC_{ist}(z_{ist}, y_{ist}, a_{ist-1})}, \tag{22}$$

Consider a strategy where firms choose the above reaction function as long as their competitors also opt for this reaction function. When a firm deviates from this reaction function however, for instance to slow down the growth rate of constrained firms, the other firms increase output and thereby punish the deviator. It is straightforward to show that such an equilibrium is a Nash equilibrium, given a condition on the discount factor β. What is analytically challenging however, is the application of equilibrium refinements, because $D^s(z_{ist}, k_{ist}, a_{ist})$ becomes non-stationary in the subgame off the equilibrium path.[17] For this reason, I focus on the equilibria containing the reaction functions in Assumption 1. An important advantage of this approach, is that it allows for an analytical solution of the comparative statics across steady states with different M_s. Additionally, the analysis remains robust to alternative assumptions on market structure, as the results will hold for both Cournot and Bertrand competition.

In each period, there is a unique y_{ist} that solves the reaction function in (22), conditional on a firm's productivity and other firms' output choices, since the left-hand side and the right-hand side of equation (22) are monotonically increasing and decreasing, respectively. For identical mathematical reasons as why the equilibrium under no financial constraints was unique, the equilibrium outcomes for any given period continue to be unique in this setting, given the above reaction functions. The intuition for the unique equilibrium is that each firm's output choice is a monotonic function of the sectoral price index $P_{st}(\mathbf{y_{ist}})$, which in turn is a concave function of each firm's output, as explained in Corchon (1994).

[17] In settings with firm heterogeneity, subgame or Markov perfect equilibria in oligopolistics industries are often solved numerically instead of analytically. See e.g. Doraszelski and Pakes (2007) for an overview.

2.5 Steady state distribution of capital

What will the distribution of input levels look like in steady state equilibrium? First, a firm's capital depends on whether it is unconstrained or constrained. When it is constrained, it grows its capital by setting it at k^c_{ist}. For unconstrained firms, as derived in Section 2.3, high-productivity firms have a higher capital level:

$$k_{sH} > k_{sL},$$

where the subscripts H, L refer to optimal choices in steady state by the unconstrained high- and low-productivity firms respectively.[18]

In the steady state equilibrium with infinitely lived firms, these capital choices will imply that either firms are at their unconstrained capital level, or that they are growing their capital from k_{sL} to k_{sH}, in the manner described by the following lemma:

Lemma 2. *In steady state, the joint distribution of capital and productivity within a sector is as follows:*

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

To see why this lemma holds, start by supposing that a firm is currently unconstrained at k_{sL}, and receives a positive productivity shock and knows that in the next period, it will be at high productivity z_{sH}. There are then two possibilities. One is that the firm is immediately unconstrained and able to set k_{sH}. The other possibility is that the firm is constrained, and in that case it will start growing its capital level by setting it at k^c_{ist+1}. Moreover, for any future period τ after its most recent positive productivity shock where the firm is still constrained, the firm will set its capital at k^c_{ist}. Then, consider firms that are at low productivity z_{sL} for the current period, which should be unconstrained at k_{sL} according to the Lemma. Suppose they were not at k_{sL}, then these firms would all be growing their capital to the next period, even if they remain at z_{sL}, and this will lead to a contradiction with $H^s(z, k, l)$ being stable over time.[19]

2.6 Comparative statics on the degree of competition

To start the analysis of comparative statics across industry equilibria, I consider industry equilibria under two different values for the number of firms: $M_s \neq M'_s$. The equilibrium values

[18] The fact that the minimum markup $\sigma/(\sigma - 1)$ is above unity implies that firms can afford an investment rate at least as high as the depreciation rate, since it implies that a return at least as high as r^k is available for each unit of capital.

[19] The stability of $H^s(z, k, l)$ requires that when *all* the z_{sL} firms are growing their capital, there should be other firms, which were z_{sH} in the previous period, that take their spot in the $H^s(z, k, l)$ distribution, with a capital level below k_{sL}. This is inconsistent with being in steady state. Firms have a wealth level either above or below k_{sL}. If it is above, then when they experience a negative productivity shock, they will choose exactly k_{sL} in the next period. If their wealth level is below k_{sL}, then we can again ask who will take their spot in the $H^s(z, k, l)$ distribution. We have an inductive step then here, with a requirement for more deeply constrained firms at each step. Since firms with higher wealth will never return to wealth below k_{sL} and given that firms are infinitely lived, this inductive step leads to a contradiction with $H^s(z, k, l)$ being stable.

under M_s' are denoted with a prime. Comparing these equilibria, I aim to examine how capital growth for constrained firms in each "bin" τ, as well as markup levels for both constrained and unconstrained firms behave as a function of the number of firms in the industry. Initially, I will be agnostic about whether $M_s > M_s'$ or not, and start instead by supposing, without loss of generality, that the low-productivity firm's market share is higher in the former equilibrium $(m_{sL} \geq m'_{sL})$, and examining the logical implications of that supposition on other firms' market shares and capital growth rates. Those logical implications, summarized in Lemma 3, will in turn imply that $M_s < M_s'$, which will allow me to conduct comparative statics across equilibria with different numbers of firms. Note that from now on, only constrained firms are denoted with the subscript τ.

Lemma 3. *If the market share of the unconstrained low-productivity firm is higher in one industry equilibrium compared to another, then the market share of the high-productivity unconstrained firms is also higher:*

$$(m_{sL} > m'_{sL}) \implies (m_{sH} > m'_{sH}). \tag{23}$$

Moreover, the capital growth rate and the market share for all constrained firms in any bin τ will be higher in the former equilibrium as well:

$$(m_{sL} > m'_{sL}) \implies \forall \tau > 0 : (G_{s\tau} > G'_{s\tau}) \wedge (m_{s\tau} > m'_{s\tau}) \tag{24}$$

To see why equation (23) holds, consider the implication of the factor demand functions (17) and (18) on the output ratio of the high- and low-productivity firm:

$$\frac{y_{sH}}{y_{sL}} = \left(\frac{z_{sH}}{z_{sL}}\right)^{\sigma} \left(\frac{\mu_{sL}}{\mu_{sH}}\right)^{\sigma} \tag{25}$$

This implies that the market shares of firm types H, L are linked, and that when m_{sL} increases, m_{sH} can only decrease if the relative markup $\frac{\mu_{sL}}{\mu_{sH}}$ increases. The proof in Appendix Section A.1 formally demonstrates that this entails a contradiction and that therefore Equation (23) holds.

Next, I demonstrate that equation (24) holds by employing a proof by induction, available in Appendix Section A.2. The proof starts by observing that if the low-productivity firms have a higher market share, their higher markup implies that they have higher capital growth immediately after a positive productivity shock: $G_{s1} > G'_{s1}$. This is because their capital growth depends on their revenue net of labor payments, per unit of capital, which can be written as the sum of profits and payments to capital:

$$\frac{\pi_{sL}}{P^F k_{sL}} = (\mu_{sL} - 1) MC_{sL} \frac{y_{sL}}{k_{sL}} + r^k.$$

Hence, since the rental rate of capital r^k is constant, variation in the capital growth depends on profits per unit of capital: $(\mu_{sL} - 1) MC_{sL} \frac{y_{sL}}{k_{sL}}$. Since marginal cost and the capital-output ratio are constant for unconstrained firms, the higher markup associated with a higher market share will entail faster capital growth.

After establishing that capital growth increases for firms in the period when they first learn about their positive productivity shock, the proof demonstrates that capital growth continues to be elevated in subsequent periods. Formally, the proof demonstrates the following inductive step:

$$((m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau})) \implies (G_{s\tau+1} > G'_{s\tau+1}).$$

I derive this inductive step by showing that revenue net of labor payments per unit of capital, $\frac{\pi_{s\tau}}{P^F k_{sL}}$, increases when $((m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau}))$. Given that $G_{s\tau} \equiv \Pi_{r=s}^{s+\tau}\left(\frac{\pi_{sv}}{P^F k_{sv}} + 1 - \delta\right)$, this implies that $(G_{s\tau+1} > G'_{s\tau+1})$. The intuition for the argument is as follows. When the low-productivity firm has a higher market share and the firm in bin τ has a higher capital growth rate $G_{s\tau}$, the firm in bin τ has the option of setting its marginal cost at the same level as the firm in bin τ under M'_s. In that case, this firm would see its markup increase, $\mu_{s\tau} > \mu'_{s\tau}$, as it faces higher demand than the firm under M'_s. A higher markup under constant marginal cost implies that $\frac{\pi_{s\tau}}{P^F k_{sL}}$ increases, and thereby capital growth increases as well. Of course, the firm can choose a different level of production at a different marginal cost. Critically though, the firm will not be worse off in terms of revenue net of labor cost if its optimal choice implies a different marginal cost, and therefore its capital growth will also be increasing in that case. In addition, the proof demonstrates that:

$$((m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau})) \implies (m_{s\tau} > m'_{s\tau}).$$

Here, the intuition is that a higher $G_{s\tau}$ reduces a firm's marginal cost, which translates into a higher market share through the reaction function in Equation (22).

Equations (23) and (24) from Lemma 3 together imply that the market shares of all types of firms, i.e. low-productivity firms L, unconstrained high-productivity firms H, and constrained firms in any bin τ, jointly increase or decrease across industry equilibria.[20] When we have $(m_{sL} > m'_{sL})$, market shares for all these types of firms increase. When market shares for all types of firms increase, it implies that there are fewer firms in the industry, and therefore:

$$(m_{sL} > m'_{sL}) \implies (M_s < M'_s).$$

The converse also holds, since $(M_s < M'_s)$ implies that the market share of at least one type of firm needs to strictly decrease. Since all market shares increase and decrease together (see footnote 20), it follows that:

$$(M_s < M'_s) \implies (m_{sL} > m'_{sL}).$$

In combination with Lemma 3, this directly implies that when the number of firms falls, the market share of all types of firms increases:

$$(M_s < M'_s) \implies ((m_{sL} > m'_{sL}) \wedge (m_{sH} > m'_{sH}) \wedge (\forall\tau : m_{s\tau} > m'_{s\tau})) \tag{26}$$

This result, combined with the monotonically increasing relationship between market shares and markups, implied by equations (5) and (22), then directly implies that markup levels for all types of firms fall as the number of firms increases. Moreover, together with equation (24) from Lemma 3, it implies that capital growth rates of constrained firms fall as well when the number of firms increases. Finally, Appendix Section A.3 demonstrates that markup dispersion also falls with the number of firms. This is intuitive, since as $M_s \to \infty$, all markups converge to $\frac{\sigma}{\sigma-1}$, the markup under monopolistic competition. Taking all the above together, this demonstrates the

[20] Naturally, the statements $(m_{sL} > m'_{sL}) \wedge (m_{sH} \leq m'_{sH})$, and $(m_{sL} > m'_{sL}) \wedge (m_{s\tau} \leq m'_{s\tau})$, for any $\tau > 0$, are in contradiction with respectively equation (23) and Equation (24).

following Proposition:

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

- *Markup levels fall with M_s:*

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

- *Markup dispersion falls with M_s:*

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

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

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

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

These results also have important welfare implications, in particular for understanding the gains from taking competition to its upper limit. Recall from Lemma 1, that when firms face no capital constraints, setting $M_s \to \infty$ equalizes marginal products across firms. In contrast, Proposition 1 entails that $\frac{k_{s\tau}}{k_{sL}}$ falls with M_s, such that the wedge between the socially optimal capital ratio $\frac{\tilde{k}_{s\tau}}{\tilde{k}_{sL}} = \left(\frac{z_{sH}}{z_{sL}}\right)^{\sigma-1}$ and the actual capital ratio actually deepens.

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.[21] 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.[22] For the dereservation analysis, I use product-level dereservation data that has been generously provided by Ishani Tewari, and a complete description of this data and its construction is available in Tewari and Wilde (2017).

When analyzing the impact of competition for the full panel of plants, I will be exploiting variation across 3-digit sectors[23] and geographical units in India. The geographical regions in the

[21] 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.

[22] 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, Collard-Wexler, and O'Connell (2016), which were generously made available by Hunt Allcott.

[23] 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.

data are either states or union territories. To make the definitions of regions consistent over time, I employ the concordance provided by the Indian Statistical Office. This results in a number of 35 regions in the data.

The main plant-level variables used in the analysis are capital K_{irst}, labor L_{irst}, materials M_{irst} and revenue S_{irst}, where subscripts indicate plant i, region 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. I winsorize variables at the second and 98th percentile. Revenue, capital and materials are deflated using deflators from the Indian Handbook of Industrial Statistics. However, whenever I calculate cost shares, I use the non-deflated variables. Below, the logarithm of a variable is denoted in lower case.

4 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 use a natural experiment - India's dereservation reform - to test the theory's predictions on the impact of competition at the plant level.

4.1 Background on the dereservation reform

The dereservation reform consists of the staggered removal of the small-scale industry (SSI) reservation policy. This reservation policy mandated that only industrial undertakings below a certain investment ceiling (10 million Rupees at historical cost in 1999) were allowed to produce certain product categories.[24] In 1996, before the start of dereservation, around 1000 product categories were reserved for SSI. However in 1997, the Indian government starts with gradually removing the reservation policy, and the process of dereservation peaks between 2002 and 2008.

A detailed description of the history of reservation policies and of the implementation of dereservation is provided by Martin et al. (2017) and Tewari and Wilde (2017), and they also provide evidence for the exogenous nature of the reform. First, Tewari and Wilde (2017) demonstrate that there is considerable variation in the timing of dereservation for strongly related product categories (e.g. different types of vegetable oils). As products within these narrow product categories arguably share similar demand and supply characteristics, this limits the scope for a structural explanation of the timing of dereservation. Moreover, Martin et al. (2017) examine pre-dereservation trends, by year of dereservation, in employment, output, capital and wages, and find no evidence for any differential trends.

Which are, from an ex-ante point of view, the main possible effects of deservation? The plant-level effects of dereservation will vary depending on whether a plant was producing reserved product categories prior to dereservation or not. Incumbent plants, whose main product was reserved prior to dereservation, are affected by dereservation in two distinct ways. First, the direct effect of the removal of the investment ceiling is that incumbents are allowed to grow their capital stock. Second, there is a 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. Martin et al. (2017) provide evidence that entry

[24] 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.

indeed occurred, since 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, market shares of incumbent plants fell. 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. Now, I first confirm that the reform also led to a reduction in markups.

4.2 Event study of dereservation's impact on markups

When testing if the dereservation reform leads to lower markup levels, I measure μ_{irt}, the markup for plant i, located in region r, in year t, by using the measurement in De Loecker and Warzynski (2012). They show that the assumptions of Cobb-Douglas production functions and cost minimization, and having labor as a variable input imply that:[25]

$$\mu_{irt} = \alpha^L_{ir} \frac{VA_{irt}}{w_{irt} L_{irt}} \tag{27}$$

where $\frac{w_{irt} L_{irt}}{VA_{irt}}$ is labor's share of value added, and α^L_{ir} is the elasticity of value added with respect to labor. It is intuitive that, when holding α^L_{ir} constant, the higher is the labor share in value added, the lower is the markup.[26] Empirically, I allow for maximal cross-plant heterogeneity in α^L_{ir} by absorbing it in a plant fixed-effect. 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}.

$$\ln \mu_{irt} = \gamma_{ir} + \nu_{rt} + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{irt} + \tau] + \varepsilon_{irt} \tag{28}$$

Here, γ_{ir} is a plant fixed-effect and ν_{rt} is a region-year fixed effect. The year in which a plant's first product is dereserved is denoted by e_{irt}, and I bin up the end points and normalize $\beta_{-1} = 0$. For the purpose of this event study, I restrict attention to a balanced sample of incumbent plants which are observed at least 5 years prior and at least four years after they were dereserved. Here, an incumbent plant is any plant that produced a reserved product prior to dereservation. In this specification, as in all the following, standard errors are clustered at the plant level.

I find that on average, markups indeed fall due to the dereservation reform (see Figure 1, Panel (a)). The initial decline is modest, but eventually the average markup declines by 0.12 log points, and significantly so. This impact of dereservation is in line with the theoretical prediction that markup levels fall when competition increases, and the economic magnitude of the impact of the reform is substantial. Note also that there is no evidence of a downward trend in markups prior to dereservation, since the estimates of β_τ for $\tau < -1$ are all not significantly different from zero.

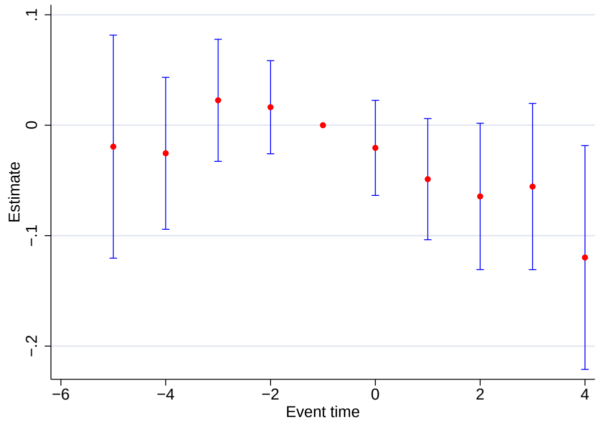
In addition to leading to a decrease in the average markup, dereservation also reduces markup dispersion. Recall that the model predicts that as the degree of competition increases, all markups

[25] Given a Cobb-Douglas production function, one can write the Lagrangian of the cost-minimization problem with variable labor input and a predetermined capital level as $\min_{l_{irt}} \mathcal{L}_{irt} = w_{irt} l_{irt} + \lambda_{irt}(Y_{irt} - a_{irt} k_{irt}^{\alpha^K_{ir}} l_{it}^{\alpha^L_{ir}})$. Which implies the first-order condition: $\frac{\partial \mathcal{L}_{irt}}{\partial l_{irt}} = w_{irt} - \lambda_{irt} \alpha^L_{ir} \frac{a_{irt} k_{irt}^{\alpha^K_{ir}} l_{it}^{\alpha^L_{ir}}}{l_{it}} = 0$, and therefore $\frac{p_{irt}}{\lambda_{irt}} = \alpha^L_{ir} \frac{p_{irt} a_{irt} k_{irt}^{\alpha^K_{ir}} l_{it}^{\alpha^L_{ir}}}{w_{irt} l_{irt}}$. Since λ_{irt} is the marginal cost of output, $\frac{p_{irt}}{\lambda_{irt}} = \mu_{irt} = \alpha^L_{ir} \frac{p_{irt} y_{irt}}{w_{irt} l_{irt}}$, and above I measure $p_{irt} y_{irt}$ in value-added terms.

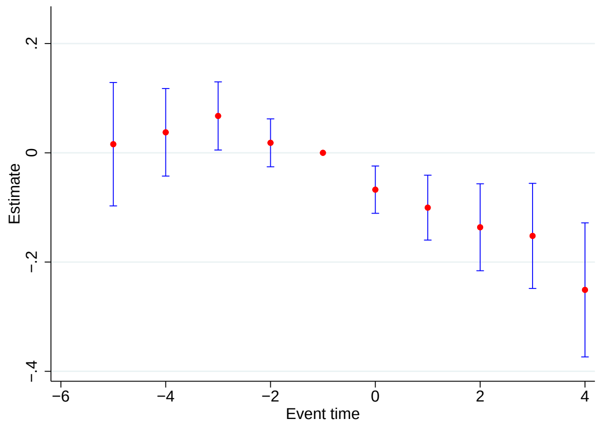
[26] This result is often used in macroeconomics to measure markups, see e.g. Nekarda and Ramey (2013) for an overview and discussion.

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

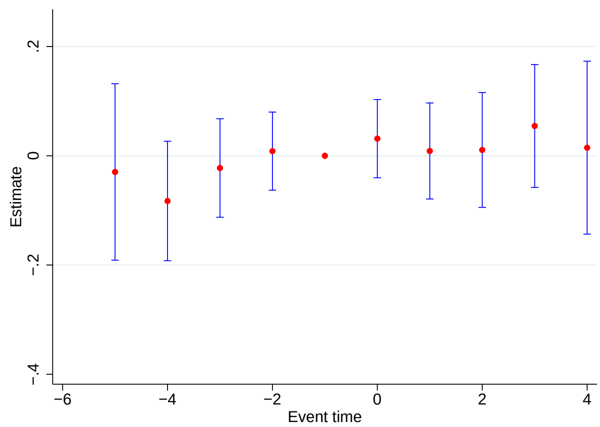
(a) Full sample



(b) Plants with above-median initial markup



(c) 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: $\ln \mu_{irt} = \gamma_{ir} + \nu_{rt} + \sum_{\tau=-5}^{4} \beta_\tau 1[t = e_{irt} + \tau] + \varepsilon_{irt}$, where γ_{ir} is a plant fixed-effect and ν_{rt} is a region-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 full sample of incumbent plants. Panel (b) displays results for plants with initial markups weakly above the median markup, and Panel (c) 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.

converge to a lower bound. To test this prediction, I split the set of incumbent plants into two subsets, depending on whether before dereservation a plants' markup is above or below the median markup. I find that plants who have higher markups in the periods before dereservation, exhibit a stronger average decline in their markup after dereservation, as predicted by the theory. More specifically, for the subset of plants with below-median initial markup, I cannot reject that dereservation has no effect on markup levels. In contrast, plants with above-median initial markups, experience an immediate and significant decline in the markup at the time of dereservation. This decline continues in the subsequent periods, reaching a reduction of 0.25 log points after 4 years (see Figure 1, Panel (b)). In Panel (b), the decline in markups for this group appears to already start before plants' first products are dereserved, since there is a decline in markups during the two years right before dereservation. Since there appears to be no downward trend in the earliest three pre-periods, this may be due to anticipation effects. After all, the final phase of the government's decision process about the dereservation of a product category involved consultation with stakeholders, which may have made it possible for both incumbent or entrant plants to anticipate dereservation.

4.3 Dereservation and capital misallocation

After documenting how the dereservation reform leads to a fall in markup levels, I now examine the reform's impact on capital convergence. In the model, firms optimally choose to grow their capital stock in response to positive productivity shocks until they reach their optimal, unconstrained level of capital. The empirical challenge is that optimal capital is unobserved. Interestingly, while a positive productivity shock leads to a first-order increase in the optimal level of capital, the change in optimal marginal revenue product of capital (MRPK) is second order. This makes it possible to find valid proxies for optimal MRPK. For this reason, and inspired by Asker et al. (2014) - henceforth ACWDL, I focus on convergence in MRPK in my empirical analysis of capital convergence. When a firm is financially constrained, its actual $MRPK_{it}$ will be above its optimal, unconstrained MRPK, denoted by $MRPK^*_{it}$. Since $MRPK_{it}$ is a strictly monotonic function of k_{it}, and capital convergence in the model slows down with M_s, convergence in terms of MRPK also slows down with M_s.

As in ACWDL, I measure MRPK in logs after assuming Cobb Douglas production functions:

$$MRPK_{it} = \ln \alpha_i^K + s_{it} - k_{it}, \tag{29}$$

where α_i^K is the output elasticity of capital. To allow α_i^K to vary across plants, I absorb this output elasticity in plant-level fixed effects in the regressions. Hence, within-plant variation in MRPK will be driven by the log difference between revenue and capital. To examine if MRPK convergence slows down due to dereservation, I use the following autoregression framework:

$$\begin{aligned} MRPK_{it} = \gamma_i + \nu_t + \beta_1 Deres_{it-1} + \rho_0 MRPK_{it-1} \\ + \rho_1 MRPK_{it-1} * Deres_{it-1} + \beta_2 \ln age_{it} + \varepsilon_{irt} \end{aligned} \tag{30}$$

where γ_i and ν_t are plant and year fixed effects respectively, and $Deres_{it-1}$ indicates if a first product of plant i has been dereserved in period $t-1$ or earlier. I estimate equation (30) on a sample with only incumbents, but also on the full sample of plants. In the latter setup, I can control for economic shocks at the region-sector-year level, which is impossible when restricting the sample to incumbents, due to collinearity issues.[27]

Table 1: Dereservation and MRPK convergence

	$MRPK_{it}$ - Value Added (VA)		$MRPK_{it}$ - Gross Revenue (GR)	
	(1)	(2)	(3)	(4)
$Deres_{it-1}$	0.191	0.206	0.125	0.186
	(0.054)	(0.054)	(0.036)	(0.037)
$MRPK_{it-1}(VA)$	0.314	0.224		
	(0.009)	(0.007)		
$MRPK_{it-1}(VA) * Deres_{it-1}$	0.064	0.044		
	(0.013)	(0.013)		
$MRPK_{it-1}(GR)$			0.427	0.350
			(0.011)	(0.008)
$MRPK_{it-1}(GR) * Deres_{it-1}$			0.063	0.058
			(0.012)	(0.012)
Plant Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	–	Yes	–
State-sector-year Fixed Effects	No	Yes	No	Yes
Observations	62442	174776	69949	204973

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 (30) 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 (31) on the full sample. All specifications control for the logarithm of a plant's age. Standard errors are clustered at the plant level.

The main coefficient of interest is ρ_1, which estimates how the speed of convergence to $MRPK^*_{it}$ changes after 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, the average plant experiences a delayed adjustment to $MRPK^*_{it}$. Crucially, when $\rho_0 > 0$, then $\rho_1 > 0$ indicates that the speed of MRPK convergence slows down after dereservation. In equation (30), I proxy for $MRPK^*_{it}$ with $\gamma_i + \nu_t + \beta_1 Deres_{it-1} + \beta_2 \ln age_{it}$, though the findings on ρ_1 are robust to the exact choice of proxy.

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

$$\begin{aligned} MRPK_{irst} = &\gamma_{irs} + \nu_{rst} + \beta_1 Deres_{irst-1} + \beta_2 Deres_{irst-1} * entrant_{irs} \\ &+ \rho_0 MRPK_{irst-1} + \rho_1 MRPK_{irst-1} * incumb_{irs} + \rho_2 MRPK_{irst-1} * entrant_{irs} \\ &+ \rho_3 MRPK_{irst-1} * Deres_{irst-1} + \rho_4 MRPK_{irst-1} * Deres_{irst-1} * entrant_{irs} \\ &+ \beta_3 X_{irst} + \varepsilon_{irst} \end{aligned} \tag{31}$$

Here, γ_{irs} is a plant fixed-effect, $entrant_{irs}$ and $incumb_{irs}$ are indicators for plant i being entrants or incumbents, and ν_{rst} is a region-sector-year fixed effect that absorbs local economic shocks. While lengthy, the above specification is still intuitive. The top row is a standard difference-in-difference framework, where I allow for different MRPK 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.

The empirical results are in line with the theoretical predictions of the model (see Table 1). First, I find that there is indeed convergence to $MRPK^*_{it}$, since ρ_0 is significantly below 1, but this convergence is not immediate as ρ_0 is also significantly above 0. This is consistent with the speed of convergence being limited by the presence of financial constraints. The point estimates for ρ_0 are 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.[28] Most importantly, all coefficients on the interaction of dereservation with $MRPK_{it-1}$ are positive, as predicted by the theory, and 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 half-life of the autoregressive process by respectively 19% and 14%.[29]

One reason why measured convergence is quite fast, is the downward bias on autoregression coefficients described by Nickell (1981). My primary objective here is to provide qualitative support for the model's predictions. It is therefore important to note that the downward bias on ρ_1, works against finding evidence for dereservation slowing down MRPK convergence.

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 the steady state assumption is more plausible.

5 Competition and MRPK convergence in the full panel

In the analysis of the predictions of the model in the full sample, I focus on MRPK convergence for two reasons. First, the prediction on capital convergence is my model's most novel one, while the predictions on markup levels and dispersion have been examined in previous research (e.g. Peters (2016); Schaumans and Verboven (2015)). Second, my inverse measure of competition is the median markup in a market, which does not allow me to examine markup misallocation. My competition measure is consistent with the model however, since in the model there is a monotonic relationship between the number of firms, which govern the degree of competition, and the first moments of the markup distribution. From the point of view of the model, an alternative measure of competition could have been the number of firms. Note however that what matters for the degree of competition is not only the number of firms, but also market size, which depends on sectoral expenditure shares and income per capita, among others. The median markup incorporates these factors directly.

As my inverse measure of competition, I use the median markup at the region-sector-year level. This competition measure is arguably exogenous from the plant's point of view, and it therefore allows me to examine the causal link between competition and MRPK convergence at the plant-level. Specifically I use $Median_{rst}[\ln \mu_{irst}]$, with:

$$\mu_{irst} = \alpha_s^L \frac{VA_{irst}}{w_{irst} L_{irst}}$$

[28] 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.

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

This markup measure is identical to equation (27), except that the elasticity α_s^L is now measured as a cost share at the sector level. Because the median markup will only enter in interactions terms in the specification below, I need to demean $Median_{rst-1}[\ln \mu_{irst-1}]$ and I also normalize it to standard deviation units. To avoid results being driven by the measurement of α_s^L, demeaning happens within sectors. Next, to ensure that the median markup is plausibly exogenous to the individual plant, I restrict the sample to cases where at least 7 plants are observed in a given region-sector-year.

To implement the empirical test on MRPK convergence, I update the autoregression framework from specification (30) in the following way:

$$\begin{aligned} MRPK_{irst} = \alpha_{irs} + \gamma_{rst} + \beta \ln age_{irst} + \rho_0 MRPK_{irst-1} \\ + \rho_1 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] + \varepsilon_{irst} \end{aligned} \tag{32}$$

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

5.1 Heterogeneity along financial dependence

The above tests on MRPK convergence have all implicitly assumed that the average plant in the sample is financially constrained. However, there is empirical heterogeneity in the degree to which plants are financially constrained, which I can leverage to further corroborate the mechanism driving the link between competition and MRPK convergence. The motivating idea is that in sectors with higher levels of financial dependence, measured as $Fin\ Dep_s$, changes in the level of sector-level competition have a stronger impact on the rate of MRPK convergence.

I employ the standard Rajan and Zingales (1998) measure of sectoral financial dependence:

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

based on data for US sectors over the 1980's.[30] 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 economies with less developed financial markets, such as India, financial constraints become especially binding in sectors with high levels of $Fin\ Dep_s$.

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:

[30] I use the original Rajan and Zingales (1998) measures of financial dependence for ISIC Rev.2 sector definitions, except that I trim the financial dependence measure such that $Fin\ Dep_s \geq 0$ to ensure a clean identification of the effect of competition in the triple interaction term in specification (33). The ISIC Rev.2 sector definitions match closely with India's NIC 1987 sector definitions. The concordance between ISIC Rev.2 and NIC 1987 is provided by the Indian Statistical Office.

$$MRPK_{irst} = \alpha_{irs} + \gamma_{rst} + \beta \ln age_{irst} + \rho_0 MRPK_{irst-1} + \rho_1 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] + \rho_2 MRPK_{irst-1} * Fin\ Dep_s + \rho_3 MRPK_{irst-1} * Median_{rst-1}[\ln \mu_{irst-1}] * Fin\ Dep_s + \varepsilon_{irst} \tag{33}$$

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.

5.2 Estimation results

The results for MRPK convergence in the full panel (see Table 2) 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, for baseline specification (32), the speed of convergence always increases with the median markup. Specifically, the coefficient on ρ_1 is always negative and strongly statistically significant(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 median markup by two standard deviations, decreases the half-life of MRPK convergence by 7%.[31]

The results for heterogeneity along financial dependence are also in line with expectations (see columns 3,4,7,8). First 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. More importantly, 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. These results imply that the median markup amplifies MRPK convergence more in sectors with higher financial dependence. Consider for instance the industry producing electric machinery, which has a relatively high level of financial dependence at $FinDep_s = 0.77$. For this sector, an increase in the median markup by two standard deviations, decreases the half-life of MRPK convergence by 7.8%, according to column 3.

6 Additional evidence from undercapitalized young plants

So far, I focused on MRPK convergence to test the effect of competition on capital convergence. The advantage of examining MRPK convergence is that any plant optimally converges to $MRPK^*_{it}$. Hence, the tests on MRPK convergence are valid in general. Still, the empirical measurement of MRPK convergence is based on the assumption of Cobb-Douglas production functions, and necessitates the use of an autoregression framework. I now complement the evidence from MRPK convergence with more transparent reduced-form evidence on competition's effect on capital growth for young plants. The data suggests that these young plants are undercapitalized, as a

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

Table 2: Competition and Speed of MRPK Convergence

	$MRPK_{irst}$ (Value added (VA))				$MRPK_{irst}$ (Gross Revenue (GR))			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$MRPK_{irst-1}(VA)$	0.266	0.228	0.244	0.196				
	(0.006)	(0.006)	(0.007)	(0.007)				
$MRPK_{irst-1}$ (VA) * $Median_{rst-1}[\ln \mu_{irst-1}]$	-0.008	-0.012	-0.002	-0.006				
	(0.001)	(0.003)	(0.002)	(0.005)				
$MRPK_{irst-1}$ (VA) * $Fin\ Dep_s$			0.035	0.031				
			(0.008)	(0.009)				
$MRPK_{irst-1}$ (VA) * $Median_{rst}[\ln(\mu_{irst})] * Fin\ Dep_s$			-0.016	0.003				
			(0.004)	(0.008)				
$MRPK_{irst-1}(GR)$					0.397	0.382	0.380	0.359
					(0.007)	(0.006)	(0.008)	(0.008)
$MRPK_{irst-1}$ (GR) * $Median_{rst-1}[\ln \mu_{irst-1}]$					-0.009	-0.007	-0.004	0.004
					(0.002)	(0.003)	(0.002)	(0.004)
$MRPK_{irst-1}$ (GR) * $Fin\ Dep_s$							0.019	0.013
							(0.009)	(0.010)
$MRPK_{irst-1}$ (GR)* $Median_{rst}[\ln(\mu_{irst}] * Fin\ Dep_s$							-0.012	-0.022
							(0.005)	(0.008)
Plant Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	–	Yes	–	Yes	–	Yes	–
State-sector-year Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	177350	145313	144592	114921	216043	180286	175663	142461

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 3-digit sectors and measured in standard deviation units. To ensure that the median markup is plausibly exogenous to the individual plant, the sample is restricted to region-sector-years consisting of at least 7 plants.

stylized fact in the existing literature is that young plants exhibit higher capital growth rates than older plants (see e.g. Evans (1987); Geurts and Van Biesebroeck (2016); Haltiwanger, Jarmin, and Miranda (2013)). As I show below, this stylized fact is also present in Indian manufacturing.

To motivate the empirical analysis of capital growth theoretically, I present a model in Appendix C where new, undercapitalized firms are born each period. 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 optimal markups and thereby slows down internally financed capital growth for young firms. This is the prediction I take to the data

I measure capital growth as:

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

Here, capital is measured as the book value of assets, which is observed both at the start of year t, and at the end. The latter value is used as measure for k_{irst+1}. By measuring capital growth this way, it is not necessary to observe plants in previous years, which increases the sample size.

Dereservation reform To examine the effect of dereservation on capital growth for young plants, I focus on a sample of incumbent plants where at least 20% of their revenue share came from a reserved product category at any time prior to dereservation. I then run the following difference-in-difference specification:

$$\begin{aligned} g(k_{irst}) = \alpha_{irs} + \gamma_{rst} + \beta_1 young_{irst} + \beta_2 Deres_{irst-1} \\ + \beta_3 Deres_{irst-1} * young_{irst} + \beta_4 \ln age_{irst} + \varepsilon_{irst}, \end{aligned} \tag{34}$$

where α_{irs} is a plant fixed effect, and γ_{rst} a state-sector-year fixed effect. I consider two different measures for $young_{irst}$, namely $[-\ln(age_{irst})]$ and the indicator variable $1(age_{irst} \leq 5)$. The prediction is that the increase in competition due to dereservation leads to slower capital growth for young plants, namely $\beta_3 < 0$. I estimate specification (34) both with and without the plant-level fixed effects, as it is ambiguous what the optimal approach is. Plant 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. Hence, capital growth should not have a stable trend for a plant.[32]

Table 3 demonstrates that dereservation has a significantly negative impact on the capital growth for young plants. The magnitude of this impact is especially strong 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 0.078 log points.

Full panel I also examine the impact of the median markup on young plants' capital growth in the full panel. To this end, I update the regression analysis to the following specification, and predict that $\beta_2 > 0$.

[32]Given that capital growth is observed within a given year, this specification allows estimation on a larger sample than the MRPK analysis for the dereservation reform. Collinearity issues will prove not to be an issue, which is why I present results only for the sample of incumbents. Results for the full sample are comparable to those for the sample of incumbents.

Table 3: Dereservation and capital growth for young plants

	Capital growth $g(k)_{it}$			
	(1)	(2)	(3)	(4)
$Deres_{it-1}$	-0.038	-0.047	0.011	0.002
	(0.024)	(0.025)	(0.010)	(0.010)
$Deres_{it} * [-\ln(age_{it})]$	-0.017	-0.015		
	(0.008)	(0.008)		
$Deres_{it-1} * 1(age_{it} \leq 5)$			-0.040	-0.078
			(0.018)	(0.022)
$[-\ln(age_{it})]$	0.017	0.021		
	(0.003)	(0.005)		
$1(age_{it} \leq 5)$			0.049	0.034
			(0.006)	(0.008)
State-sector-year Fixed Effects	Yes	Yes	Yes	Yes
Plant Fixed Effects	No	Yes	No	Yes
Observations	107919	99488	109236	100843

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.

$$g(k_{irst}) = \alpha_{irs} + \gamma_{rst} + \beta_1 young_{irst} + \beta_2 Median_{rst}[\ln(\mu_{irst-1})] * young_{irst} + \varepsilon_{irst} \quad (35)$$

Finally, 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 ($\beta_3 > 0$).

$$\begin{aligned} g(k_{irst}) =& \alpha_{rst} + \beta_1 young_{irst} + \beta_2 Median_{rst}[\ln(\mu_{irst-1})] * young_{irst} \\ &+ \beta_3 Median_{rst}[\ln(\mu_{irst-1})] * young_{irst} * Fin\ Dep_s + \varepsilon_{irst} \end{aligned} \quad (36)$$

The estimation results are again generally in line with the theoretical predictions (see Table 4). 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.

7 Conclusion

Misallocation of resources is a pervasive challenge throughout the developing world. While the exact contribution of misallocation to cross-country differences in aggregate productivity is debatable, it seems uncontroversial that shifting resources to firms with higher marginal products would be particularly beneficial for developing economies. When quantifying the benefits of re-

Table 4: Competition and capital growth of young plants

	Capital growth $g(k)_{irst}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Median_{rst}[\ln(\mu_{irst})] * [-\ln(age_{irst})]$	0.003	0.003			0.001	0.002		
	(0.001)	(0.002)			(0.002)	(0.003)		
$Median_{rst}[\ln(\mu_{irst-1})] * 1(age_{irst} \leq 5)$			0.007	-0.005			0.005	-0.005
			(0.002)	(0.004)			(0.004)	(0.005)
$Median_{rst}[\ln(\mu_{irst})] * [-\ln(age_{irst})] * Fin\ Dep_s$					0.010	0.006		
					(0.003)	(0.006)		
$Median_{rst}[\ln(\mu_{irst})] * 1(age_{irst} \leq 5) * Fin\ Dep_s$							0.019	0.006
							(0.008)	(0.012)
$-\ln(age_{irst})$	0.013	0.018			0.009	0.012		
	(0.001)	(0.003)			(0.002)	(0.003)		
$1(age_{irst} \leq 5)$			0.058	0.022			0.048	0.014
			(0.002)	(0.003)			(0.003)	(0.005)
$-\ln(age_{irst}) * Fin\ Dep_s$					0.011	0.013		
					(0.003)	(0.006)		
$1(age_{irst} \leq 5) * Fin\ Dep_s$							0.024	0.005
							(0.006)	(0.009)
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	594124	486085	607068	497116	503482	406177	514896	415778

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 (35), while columns 5-8 show results for specification (36). To ensure that the median markup is plausibly exogenous to the individual plant, the sample is restricted to region-sector-years consisting of at least 7 plants.

ducing misallocation, a typical approach is to focus on one particular friction, for instance market power, financial constraints, or a particular government intervention, and ask how the removal of this friction could contribute to aggregate productivity. This paper shows that this type of approach can lead to misleading conclusions.

I have examined how the interplay of competition with financial constraints affects misallocation, and found that increased competition reduces markup misallocation, but amplifies capital misallocation. These countervailing forces make the impact of intensified competition on misallocation ambiguous. I have also documented a range of empirical results for India's manufacturing sector, all indicating that increased competition indeed slows down firms' convergence to their optimal capital level. Hence, the empirical relevance of competition's negative impact on capital convergence is robustly demonstrated.

Deriving precise policy implications is outside the scope of this paper. Nevertheless, the analysis suggests that reaping the full gains of pro-competitive reforms depends very much on the precise implementation of these reforms. More precisely, it is essential that reforms do not stifle the growth of high-potential firms. In thinking about sequencing of reforms, the take-away could be to first optimize financial access for firms, before enhancing competition in the real economy.

References

Acemoglu, D. and M. K. Jensen (2013). Aggregate comparative statics. *Games and Economic Behavior 81*, 27–49.

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

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and Innovation: An Inverted-U Relationship. *The Quarterly Journal of Economics*, 701–728.

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

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

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

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

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

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2017). The elusive pro-competitive effects of trade. *The Review of Economic Studies (forthcoming)*.

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

Asturias, J., M. García-Santana, and R. Ramos (2018). Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India. *Journal of the European Economic Association (forthcoming)*.

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

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

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

Bhaduri, S. N. (2005). Investment, financial constraints and financial liberalization: Some stylized facts from a developing economy, india. *Journal of Asian Economics 16*(4), 704–718.

Bils, M., P. J. Klenow, and C. Ruane (2017). Misallocation or mismeasurement? *Working Paper*.

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

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

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

Brooks, W. J., J. P. Kaboski, and Y. A. Li (2016). Growth policy, agglomeration, and (the lack of) competition. *NBER working paper*.

Buera, F. and J. Nicolini (2017). Liquidity traps and monetary policy: Managing a credit crunch. *working paper*.

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

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

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

Corchon, L. C. (1994). Comparative statics for aggregative games the strong concavity case. *Mathematical Social Sciences 28*(3), 151–165.

De Loecker, J. and J. Eeckhout (2017). The rise of market power and the macroeconomic implications. Technical report.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

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

Dhingra, S. and J. Morrow (2016). Monopolistic competition and optimum product diversity under firm heterogeneity. *working paper*.

Doraszelski, U. and A. Pakes (2007). A framework for applied dynamic analysis in io. *Handbook of industrial organization 3*, 1887–1966.

Edmond, C., V. Midrigan, and D. Y. Xu (2015). Competition, markups, and the gains from international trade. *American Economic Review 105*(10), 3183–3221.

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

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

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

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

Geurts, K. and J. Van Biesebroeck (2016). Firm creation and post-entry dynamics of de novo entrants. *International Journal of Industrial Organization 49*, 59–104.

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

Gopinath, G., Ş. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017). Capital allocation and productivity in south europe. *The Quarterly Journal of Economics 132*(4), 1915–1967.

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

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annu. Rev. Econ. 6*(1), 735–770.

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

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

Itskhoki, O. and B. Moll (2018). 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.

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

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

Krusell, P., T. Mukoyama, and A. A. Smith Jr (2011). Asset prices in a huggett economy. *Journal of Economic Theory 146*(3), 812–844.

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

Macchiavello, R. and A. Morjaria (2015). Competition and relational contracts: Evidence from Rwanda's mills. *working paper*.

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

Martin, L. A., S. Nataraj, and A. E. Harrison (2017). In with the big, out with the small: Removing small-scale reservations in india. *American Economic Review 107*(2), 354–86.

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

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

Mrázová, M., J. P. Neary, and M. Parenti (2017). Sales and markup dispersion: Theory and empirics. *working paper*.

Nekarda, C. J. and V. A. Ramey (2013). The cyclical behavior of the price-cost markup. *NBER working paper*.

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

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

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

Ravn, M. O. and V. Sterk (2016). Macroeconomic fluctuations with hank & sam: An analytical approach. *working paper*.

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

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

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

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

Ventura, J. and H.-J. Voth (2015). Debt into growth: how sovereign debt accelerated the first industrial revolution. *NBER working paper*.

Werning, I. (2015). Incomplete markets and aggregate demand. *NBER Working Paper*.

Appendix A Proof for Lemma 3

A.1 Proof for Equation (23)

In this proof by contradiction, I will demonstrate that Equation (23) holds, namely $(m_{sL} > m'_{sL}) \implies (m_{sH} > m'_{sH})$. Suppose to the contrary[33] that $(m_{sL} > m'_{sL}) \wedge (m_{sH} \leq m'_{sH})$. First defining the output ratio:

$$\mathcal{G}_{sH} \equiv \left(\frac{y_{sH}}{y_{sL}}\right)^{\frac{\sigma-1}{\sigma}},$$

which from Equation (6) implies that $m_{sH} = \mathcal{G}_{sH} m_{sL}$. Hence the supposition implies that the output ratio shrinks:

$$(m_{sL} \geq m'_{sL}) \wedge (\mathcal{G}_{sH} < \mathcal{G}'_{sH}).$$

Equation (25) then entails that $(\mathcal{G}_{sH} < \mathcal{G}'_{sH}) \implies \left(\frac{\mu'_{sL}}{\mu'_{sH}} > \frac{\mu_{sL}}{\mu_{sH}}\right)$. Or alternatively, since all markups are positive:

$$(\mathcal{G}_{sH} < \mathcal{G}'_{sH}) \implies \left(\frac{\mu_{sH}}{\mu'_{sH}} > \frac{\mu_{sL}}{\mu'_{sL}}\right).$$

However, note that $(m_{sL} > m'_{sL}) \implies \left(\frac{\mu_{sL}}{\mu'_{sL}} > 1\right)$. After also observing that $\left(\frac{\mu_{sH}}{\mu'_{sH}} > 1\right) \iff (m_{sH} > m'_{sH})$, the combination of the previous results yields:

$$((m_{sL} > m'_{sL}) \wedge (\mathcal{G}_{sH} < \mathcal{G}'_{sH})) \implies (m_{sH} > m'_{sH}).$$

This entails a contradiction with the supposition. Hence, its opposite must be true, which proves the statement in Equation (23).

A.2 Proof for Equation (24)

I start by demonstrating the first component of the implication in Equation (24), namely:

$$(m_{sL} > m'_{sL}) \implies \forall \tau > 0 : (G_{s\tau} > G'_{s\tau}) \tag{37}$$

To show this, I start from the expression for capital growth, $k_{s\tau+1} = (1-\delta)k_{s\tau} + \frac{\pi_{s\tau}}{PF}$, which implies:

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

$$G_{s\tau+1} \equiv \frac{k_{s\tau+1}}{k_{sL}} = (1-\delta)G_{s\tau} + \frac{\pi_{s\tau}}{P^F k_{sL}} \tag{38}$$

I then demonstrate Equation (37) by induction. In a first step, I show that $(m_{sL} > m'_{sL}) \implies (G_{s1} > G'_{s1})$, and afterwards I demonstrate the inductive step that $(m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau}) \implies (G_{s\tau+1} > G'_{s\tau+1})$. The first step and the inductive step together imply that Equation (37) holds.

Step 1 First, notice that revenue net of labor costs of the unconstrained low-productivity firm is $\pi_{sL}/P^F = (\mu_{sL}-1)y_{sL}MC_{sL} + r^k k_{sL}$, where the first term are profits, and the second term is "payments to capital." Hence

$$\frac{\pi_{sL}}{p^F k_{sL}} = (\mu_{sL}-1)MC_{sL}\frac{y_{sL}}{k_{sL}} + r^k.$$

Recall that $(m_{sL} > m'_{sL}) \implies (\mu_{sL} > \mu'_{sL})$, while $MC_{sL} = MC'_{sL}$ (see Equation (19)). In addition, Equations (17) and (18) imply that the output to capital ratio is constant at $\frac{y_{sL}}{k_{sL}} = \frac{P^F}{w} r^k \frac{(1-\alpha)}{\alpha}$, which implies $\frac{y_{sL}}{k_{sL}} = \frac{y'_{sL}}{k'_{sL}}$. Taken together, $(m_{sL} > m'_{sL}) \implies \left(\frac{\pi_{sL}}{P^F y_{sL}} > \frac{\pi'_{sL}}{P^F y'_{sL}}\right)$. Since $G_{s1} \equiv \frac{k_{s1}}{k_{sL}} = (1-\delta) + \frac{\pi_{sL}}{P^F y_{sL}}$, I obtain that

$$(m_{sL} > m'_{sL}) \implies (G_{s1} > G'_{s1}).$$

Inductive Step I now demonstrate that the inductive step holds, i.e.:

$$(m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau}) \implies (G_{s\tau+1} > G'_{s\tau+1}).$$

Rewriting Equation (38), note that $G_{s\tau+1} = (1-\delta)G_{s\tau} + \frac{\pi_{s\tau}}{P^F k_{sL}}$. If $G_{s\tau} > G'_{s\tau}$, then a sufficient condition to have $(G_{s\tau+1} > G'_{s\tau+1})$ is therefore that

$$\frac{\pi_{s\tau}}{P^F k_{sL}} \geq \frac{\pi'_{s\tau}}{P^F k'_{sL}}.$$

It is relatively straightforward to show that:[34]

$$\frac{\pi_{s\tau}}{P^F k_{sL}} = (\mu_{s\tau} - (1-\alpha))\frac{l_{s\tau}}{k_{s\tau}} G_{s\tau} \frac{w}{P^F(1-\alpha)}, \tag{39}$$

and in what follows I examine how $\frac{\pi_{s\tau}}{P^F k_{sL}}$ is related to m_{sL} and $G_{s\tau}$.

In order to demonstrate the inductive step, I first show that:

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

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

Given the Cobb-Douglas production function, we have that total labor costs for any given quantity $\bar{y}_{s\tau}$ are $TLC(\bar{y}_{s\tau}) = \frac{w}{P^F} l(\bar{y}_{s\tau})$. For constrained firms, setting $\bar{y}_{s\tau}$ directly implies setting the amount of labor in the following function: $l(\bar{y}_{s\tau}) = \left(\frac{\bar{y}_{s\tau}}{z_{sH}k^{\alpha}_{s\tau}}\right)^{\frac{1}{1-\alpha}}$, such that $TLC(\bar{y}_{s\tau}) = \frac{w}{P^F}\left(\frac{\bar{y}_{s\tau}}{z_{sH}k^{\alpha}_{s\tau}}\right)^{\frac{1}{1-\alpha}}$. Hence, $ALC_{s\tau}(\bar{y}_{s\tau}) = \frac{w}{P^F}\left(\frac{\bar{y}^{\alpha}_{s\tau}}{z_{sH}k^{\alpha}_{s\tau}}\right)^{\frac{1}{1-\alpha}}$, and given these firms' marginal cost is as in Equation (21), $\frac{ALC_{s\tau}}{MC_{s\tau}} = (1-\alpha)$. Together with $\frac{\bar{y}_{s\tau}}{z_{sH}k^{\alpha}_{s\tau}} = l^{1-\alpha}_{s\tau}$, this implies that:

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

$$((m_{sL} > m'_{sL}) \wedge (G_{s\tau} = G'_{s\tau})) \implies (G_{s\tau+1} > G'_{s\tau+1}),$$

and then afterwards demonstrate that $((m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau})) \implies (G_{s\tau+1} > G'_{s\tau+1})$ holds a fortiori. Recall that $(m_{sL} > m'_{sL}) \implies (\mu_{sL} > \mu_{sL})$. I then analyze the implications of $(m_{sL} > m'_{sL}) \wedge (G_{s\tau} = G'_{s\tau})$ under two exhaustive cases; first $\frac{l_{s\tau}}{k_{s\tau}} = \frac{l'_{s\tau}}{k'_{s\tau}}$, and second $\frac{l_{s\tau}}{k_{s\tau}} \neq \frac{l'_{s\tau}}{k'_{s\tau}}$.

- Case (i): suppose $\frac{l_{s\tau}}{k_{s\tau}} = \frac{l'_{s\tau}}{k'_{s\tau}}$. First, recall that $MC_{sL} = MC'_{sL}$. Combined with $\mu_{sL} > \mu'_{sL}$ this implies that $p_{sL} > p'_{sL}$. Next, the inverse demand function in Equation (2.1) implies that $\frac{p_{s\tau}}{p_{sL}} = \left(\frac{q_{s\tau}}{q_{sL}}\right)^{-1/\sigma} = \left(\frac{z_{sH} l_{s\tau}}{a_L l_{sL}} G_{s\tau}\right)^{-1/\sigma}$, which I rewrite as

$$\frac{p_{s\tau}}{p_{sL}} = \left(\frac{z_{sH}}{a_L} G_{s\tau} \frac{l_{s\tau}}{k_{s\tau}} \frac{k_{sL}}{l_{sL}} G_{s\tau}\right)^{-1/\sigma}.$$

  This implies $\frac{p_{s\tau}}{p_{sL}} = \frac{p'_{s\tau}}{p'_{sL}}$, since by assumption $G_{s\tau} = G'_{s\tau}$ and $\frac{l_{s\tau}}{k_{s\tau}} = \frac{l'_{s\tau}}{k'_{s\tau}}$, while $\frac{l_{sL}}{k_{sL}} = \frac{l'_{sL}}{k'_{sL}}$ from factor demand Equations (17) and (18). Since $p_{sL} > p'_{sL}$ and $\frac{p_{s\tau}}{p_{sL}} = \frac{p'_{s\tau}}{p'_{sL}}$, I find that

$$p_{s\tau} > p'_{s\tau}.$$

  Note then finally from Equation (21) that $MC_{s\tau} = \frac{w}{z_{sH}(1-\alpha)P^F} \left(\frac{l_{s\tau}}{k_{s\tau}}\right)^{\alpha}$. So $\frac{l_{s\tau}}{k_{s\tau}} = \frac{l'_{s\tau}}{k'_{s\tau}}$ implies a constant marginal cost at $MC_{s\tau} = MC'_{s\tau}$. Taken together then, $MC_{s\tau} = MC'_{s\tau}$ and $p_{s\tau} > p'_{s\tau}$ imply that

$$\mu_{s\tau} > \mu'_{s\tau}.$$

  Therefore, using Equation (39), I find that

$$\left((m_{sL} > m'_{sL}) \wedge (G_{s\tau} = G'_{s\tau}) \wedge \left(\frac{l_{s\tau}}{k_{s\tau}} = \frac{l'_{s\tau}}{k'_{s\tau}}\right)\right) \implies \left(\frac{\pi_{s\tau}}{P^F k_{sL}} > \frac{\pi'_{s\tau}}{P^F k'_{sL}}\right). \quad (40)$$

- Case (ii): suppose the firm, following the reaction function in Equation (22), chooses optimally to have $\frac{l_{s\tau}}{k_{s\tau}} \neq \frac{l'_{s\tau}}{k'_{s\tau}}$. Since the reaction function in Equation (22) maximizes the revenue net of labor costs of the firm within period τ, and since $\frac{l_{s\tau}}{k_{s\tau}} = \frac{l'_{s\tau}}{k'_{s\tau}}$ is within the firm's choice set, setting $\frac{l_{s\tau}}{k_{s\tau}} \neq \frac{l'_{s\tau}}{k'_{s\tau}}$ implies that its revenue net of labor costs is weakly higher than under case (i), so Equation (40) continues to hold.

So far, I have assumed that $G_{s\tau} = G'_{s\tau}$. Now, consider instead $G_{s\tau} > G'_{s\tau}$. In this case, the firm with $G_{s\tau}$ can produce an identical output level as when $G_{s\tau} = G'_{s\tau}$, but with less labor input, which would lead to identical revenue at a lower labor cost. Hence, $\pi_{s\tau}$, revenue net of labor cost, is increasing. Moreover, k_{sL} is fixed, and hence, the firm's capital growth rate, $G_{s\tau+1} = (1-\delta)G_{s\tau} + \frac{\pi_{s\tau}}{P^F k_{sL}}$, will increase, and Equation (40) continues to hold. This completes the demonstration of the inductive step.

Impact on market shares Having shown that $((m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau})) \implies (G_{s\tau+1} > G'_{s\tau+1})$, what remains to be shown is that $((m_{sL} > m'_{sL}) \wedge (G_{s\tau} > G'_{s\tau})) \implies (m_{s\tau} > m'_{s\tau})$. In order to do so, start by examining $\mathcal{G}_{s\tau} \equiv \left(\frac{y_{s\tau}}{y_{sL}}\right)^{\frac{\sigma-1}{\sigma}}$, which based on the reaction function in Equation (22), is equal to:

$$\mathcal{G}_{s\mathcal{T}} \equiv \left(\frac{y_{s\mathcal{T}}}{y_{sL}}\right)^{\frac{\sigma-1}{\sigma}} = \left(\frac{MC_{sL}}{MC_{s\mathcal{T}}}\frac{\mu_{sL}}{\mu_{s\mathcal{T}}}\right)^{\sigma-1}$$

Since $m_{s\mathcal{T}} = \mathcal{G}_{s\mathcal{T}} m_{sL}$, when $m_{sL} > m'_{sL}$, a necessary condition to have $m_{s\mathcal{T}} \leq m'_{s\mathcal{T}}$ is that:

$$\left(\frac{MC_{sL}}{MC_{s\mathcal{T}}} < \frac{MC'_{sL}}{MC'_{s\mathcal{T}}}\right) \vee \left(\frac{\mu_{sL}}{\mu_{s\mathcal{T}}} < \frac{\mu'_{sL}}{\mu'_{s\mathcal{T}}}\right).$$

I now demonstrate that given $(m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}})$, $\left(\left(\frac{MC_{sL}}{MC_{s\mathcal{T}}} < \frac{MC'_{sL}}{MC'_{s\mathcal{T}}}\right) \vee \left(\frac{\mu_{sL}}{\mu_{s\mathcal{T}}} < \frac{\mu'_{sL}}{\mu'_{s\mathcal{T}}}\right)\right) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$. Hence, $(m_{s\mathcal{T}} \leq m'_{s\mathcal{T}})$ cannot hold, and therefore I will have shown what I set out to show, namely $((m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}})) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$.

- Here I show that $\left((m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}}) \wedge \left(\frac{\mu_{sL}}{\mu_{s\mathcal{T}}} < \frac{\mu'_{sL}}{\mu'_{s\mathcal{T}}}\right)\right) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$. Recall that $(m_{sL} > m'_{sL}) \implies (\frac{\mu_{sL}}{\mu'_{sL}} > 1)$, and note that given that all markups are positive: $\left(\frac{\mu_{sL}}{\mu_{s\mathcal{T}}} < \frac{\mu'_{sL}}{\mu'_{s\mathcal{T}}}\right) \implies \left(\frac{\mu_{sL}}{\mu'_{sL}} < \frac{\mu_{s\mathcal{T}}}{\mu'_{s\mathcal{T}}}\right)$. Hence, $\left((m_{sL} > m'_{sL}) \wedge \left(\frac{\mu_{sL}}{\mu_{s\mathcal{T}}} < \frac{\mu'_{sL}}{\mu'_{s\mathcal{T}}}\right)\right) \implies (1 < \frac{\mu_{s\mathcal{T}}}{\mu'_{s\mathcal{T}}})$. Since $(1 < \frac{\mu_{s\mathcal{T}}}{\mu'_{s\mathcal{T}}}) \iff (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$, and since $(G_{s\mathcal{T}} > G'_{s\mathcal{T}})$ is consistent with $(m_{s\mathcal{T}} > m'_{s\mathcal{T}})$, I obtain:

$$\left((m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}}) \wedge \left(\frac{\mu_{sL}}{\mu_{s\mathcal{T}}} < \frac{\mu'_{sL}}{\mu'_{s\mathcal{T}}}\right)\right) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$$

- Here I show that $\left((m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}}) \wedge \left(\frac{MC_{sL}}{MC_{s\mathcal{T}}} < \frac{MC'_{sL}}{MC'_{s\mathcal{T}}}\right)\right) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$. Note that

$$MC_{s\mathcal{T}} = \frac{w}{(1-\alpha)P^F}\frac{1}{z_{sH}}\left(\frac{l_{s\mathcal{T}}}{k_{s\mathcal{T}}}\right)^{\alpha}$$

Since $MC_{sL} = MC'_{sL}$, $\left(\frac{MC_{sL}}{MC_{s\mathcal{T}}} < \frac{MC'_{sL}}{MC'_{s\mathcal{T}}}\right) \implies (\frac{l_{s\mathcal{T}}}{k_{s\mathcal{T}}} > \frac{l'_{s\mathcal{T}}}{k'_{s\mathcal{T}}})$, with $\frac{l_{s\mathcal{T}}}{k_{s\mathcal{T}}} = \frac{l_{s\mathcal{T}}}{G_{s\mathcal{T}} k_{sL}}$. Hence

$$\left((G_{s\mathcal{T}} > G'_{s\mathcal{T}}) \wedge \left(\frac{MC_{sL}}{MC_{s\mathcal{T}}} < \frac{MC'_{sL}}{MC'_{s\mathcal{T}}}\right)\right) \implies \left(\frac{l_{s\mathcal{T}}}{k_{sL}} > \frac{l'_{s\mathcal{T}}}{k'_{sL}}\right).$$

Now, note that $\mathcal{G}_{s\mathcal{T}} = \frac{z_{sH}}{a_L} G^{\alpha}_{s\mathcal{T}} \left(\frac{l_{s\mathcal{T}}}{l_{sL}}\right)^{\alpha}$. From cost minimization, it follows that $l_{sL} = \left(\frac{r^k}{\alpha}\right)^{\frac{1}{1-\alpha}} k_{sL}$. Hence, $\mathcal{G}_{s\mathcal{T}} = \frac{z_{sH}}{a_L} G^{\alpha}_{s\mathcal{T}} \left(\frac{l_{s\mathcal{T}}}{\left(\frac{r^k}{\alpha}\right)^{\frac{1}{1-\alpha}} k_{sL}}\right)^{\alpha}$. Therefore, $\left((G_{s\mathcal{T}} > G'_{s\mathcal{T}}) \wedge (\frac{l_{s\mathcal{T}}}{k_{sL}} > \frac{l'_{s\mathcal{T}}}{k'_{sL}})\right) \implies (\mathcal{G}_{s\mathcal{T}} > \mathcal{G}'_{s\mathcal{T}})$. Hence:

$$\left((m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}}) \wedge \left(\frac{MC_{sL}}{MC_{s\mathcal{T}}} < \frac{MC'_{sL}}{MC'_{s\mathcal{T}}}\right)\right) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$$

Combining the previous two bullet points, given $(m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}})$, the necessary condition to have $(m_{s\mathcal{T}} \leq m'_{s\mathcal{T}})$ itself implies $(m_{s\mathcal{T}} > m'_{s\mathcal{T}})$, and therefore:

$$((m_{sL} > m'_{sL}) \wedge (G_{s\mathcal{T}} > G'_{s\mathcal{T}})) \implies (m_{s\mathcal{T}} > m'_{s\mathcal{T}})$$

A.3 Proof on Markup Dispersion

First, I examine how the markups of constrained firms behave relative to those of unconstrained firms:

$$\frac{\partial \frac{\mu_{s\mathcal{T}}}{\mu_{sL}}}{\partial M_s} = \frac{\partial \frac{\varepsilon(m_{s\mathcal{T}})\varepsilon(m_{sL}) - \varepsilon(m_{s\mathcal{T}})}{\varepsilon(m_{s\mathcal{T}})\varepsilon(m_{sL}) - \varepsilon(m_{sL})}}{\partial M_s}$$

Working out the derivative and simplifying:

$$\frac{\partial \frac{\mu_{s\mathcal{T}}}{\mu_{sL}}}{\partial M_s} = \frac{\varepsilon(m_{s\mathcal{T}})(\varepsilon(m_{s\mathcal{T}}) - 1)\frac{\partial \varepsilon(m_{sL})}{\partial m_{sL}}\frac{\partial m_{sL}}{\partial M_s} - \varepsilon(m_{sL})(\varepsilon(m_{sL}) - 1)\frac{\partial \varepsilon(m_{s\mathcal{T}})}{\partial m_{s\mathcal{T}}}\frac{\partial m_{s\mathcal{T}}}{\partial M_s}}{(\varepsilon(m_{s\mathcal{T}})\varepsilon(m_{sL}) - \varepsilon(m_{sL}))^2}$$

Plugging in the values for $\frac{\partial \varepsilon(m_{ist})}{\partial m_{ist}}$ for the Cournot demand elasticity:[35]

$$\frac{\partial \frac{\mu_{s\mathcal{T}}}{\mu_{sL}}}{\partial M_s} = \frac{(1 - \frac{1}{\sigma})\frac{\varepsilon(m_{sL})(\varepsilon(m_{sL})-1)}{\varepsilon(m_{s\mathcal{T}})^2}\frac{\partial m_{s\mathcal{T}}}{\partial M_s} - (1 - \frac{1}{\sigma})\frac{\varepsilon(m_{s\mathcal{T}})(\varepsilon(m_{s\mathcal{T}})-1)}{\varepsilon(m_{sL})^2}\frac{\partial m_{sL}}{\partial M_s}}{(\varepsilon(m_{s\mathcal{T}})\varepsilon(m_{sL}) - \varepsilon(m_{sL}))^2} < 0 \tag{41}$$

Here we have that $\frac{\varepsilon(m_{s\mathcal{T}})(\varepsilon(m_{s\mathcal{T}})-1)}{\varepsilon(m_{sL})^2} < \frac{\varepsilon(m_{sL})(\varepsilon(m_{sL})-1)}{\varepsilon(m_{s\mathcal{T}})^2}$. I then consider two cases, with first $\frac{\partial \frac{y_{s\mathcal{T}}}{y_{sL}}}{\partial M_s} > 0$, and then its opposite.

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

$$\frac{\partial \frac{\mu_{s\mathcal{T}}}{\mu_{sL}}}{\partial M_s} < 0.$$

- In case (ii), suppose $\frac{\partial \frac{y_{s\mathcal{T}}}{y_{sL}}}{\partial M_s} \leq 0$. It is then straightforward that $\frac{\partial \frac{y_{s\mathcal{T}}}{y_{sL}}}{\partial M_s} \leq 0 \iff \frac{\partial \frac{m_{s\mathcal{T}}}{m_{sL}}}{\partial M_s} \leq 0$, where $\frac{\partial \frac{m_{s\mathcal{T}}}{m_{sL}}}{\partial M_s} \leq 0 \implies \left(\frac{\partial m_{s\mathcal{T}}}{\partial M_s} m_{sL} \leq \frac{\partial m_{sL}}{\partial M_s} m_{s\mathcal{T}}\right)$. Since $m_{sL} < m_{s\mathcal{T}}$ and $\frac{\partial m_{sL}}{\partial M_s}, \frac{\partial m_{s\mathcal{T}}}{\partial M_s} < 0$ from Lemma 3, in this second case it holds that:

$$\frac{\partial \frac{y_{s\mathcal{T}}}{y_{sL}}}{\partial M_s} \implies \frac{\partial m_{s\mathcal{T}}}{\partial M_s} \leq \frac{\partial m_{sL}}{\partial M_s} < 0.$$

  Given Equation (41), this has the same implication as in the first case, namely that:

$$\frac{\partial \frac{\mu_{s\mathcal{T}}}{\mu_{sL}}}{\partial M_s} < 0.$$

The remaining question is how the relative markups of unconstrained firms behave. Here, note that $m_{sH} = \mathcal{G}_{sH} m_{sL}$, such that $\frac{\partial m_{sH}}{\partial M_s} = \mathcal{G}_{sH}\frac{\partial m_{sL}}{\partial M_s} + m_{sL}\frac{\partial \mathcal{G}_{sH}}{\partial M_s}$. Now, suppose $\frac{\partial \frac{\mu_{sH}}{\mu_{sL}}}{\partial M_s} \geq 0$. Then the factor ratios $\frac{k_{sH}}{k_{sL}} = \frac{l_{sH}}{l_{sL}} = \left(\frac{z_{sH}}{z_{sL}}\right)^{\sigma-1}\left(\frac{\mu_{sL}}{\mu_H}\right)^{\sigma}$ imply that $\frac{\partial \mathcal{G}_{sH}}{\partial M_s} < 0$. In that case, however, $\frac{\partial m_{sH}}{\partial M_s} < \frac{\partial m_{sL}}{\partial M_s}$, which from equation (41) implies that $\frac{\partial \frac{\mu_{sH}}{\mu_{sL}}}{\partial M_s} < 0$. Hence, the supposition that $\frac{\partial \frac{\mu_{sH}}{\mu_{sL}}}{\partial M_s} \geq 0$ entails a contradiction, and therefore its opposite must be true:

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

$$\frac{\partial \frac{\mu_{sH}}{\mu_{sL}}}{\partial M_s} < 0.$$

Appendix B Assumptions on the productivity volatility process

By definition, in steady state, $D^s(a', k, z)$ is stable over time for each sector, and this distribution is described by Lemma 2. Given that M_s is finite, the law of large numbers does not hold, and for a stochastic process with iid transition probabilities, $D^s(a', k, z)$ will not be exactly stable. To sidestep this issue, I do not assume that transition probabilities are iid. Instead, I make the following assumption:

Assumption. *Productivity shocks are 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.* [36]

Here, the different states x and y are defined in relation to Lemma 2. Before giving an exact definition of these states, we first need to keep track of following implications of Proposition 1.[37] Since these implications hold for any sector, we drop the subscript s.

- 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 is always strictly positive.
- Implication 2: The number of periods it takes for a high productivity firm to become unconstrained is increasing with M_s.[38] 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

Consider a sufficiently high M, $\bar{M}$, such that $\bar{M}$ will be the highest value of M considered in the comparative statics on M. Importantly, given implications 1 and 2, we have that $\forall M < \bar{M}$: $T^M \leq T^{\bar{M}}$.

Based on implication 3, the productivity volatility process will then be defined by transition probabilities across the low-productivity bin L and the high-productivity bins $T^{\bar{M}}, H^{\bar{M}}$, where $H^{\bar{M}}$ denotes the bin with all the firms that are still high-productivity beyond bin $T^{\bar{M}}$. 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.

Definition. ***The transition probabilities*** *across bins are:*

- *Probability to transition from a_L to z_{sH}, i.e. probability to transition from L to $\tau = 1$: P_{LH}. Then, $(1 - P_{LH})$ is probability to remain within L.*

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

[37] This proposition is demonstrated conditional on the steady state existing for different values of M_s, 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 exists, and that the process is identical across M_s, then Proposition 1 holds.

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

- *For firms with z_{sH}, the transition probabilities are dependent on τ. Conditional on having z_{sH} in period τ, the probability of continuing having z_{sH} is $P_{HH\tau}$.*
  - *Therefore, conditional on having z_{sH} in $\tau = 1$, the probability of still having z_{sH} in $\tau > 1$, is $\prod_{r=1}^{\tau-1} P_{HHr}$.*
- *Finally, the transition probability of moving from bin $H^{\bar{M}}$ to bin L, is P_{HL}.*

By specifying these bins and defining transition probabilities across them, I have assured that the number of firms in each bin is stable across periods.[39] To see this, denote the number of firms in L by M_L. Then, the number of firms in bin τ: $M_L P_{LH} \prod_{r=1}^{\tau-1} P_{HHr}$. Next, the number of firms in $H_{\bar{M}}^U$ can be found by setting the number of exiters from $H^{\bar{M}}$ equal to the number of entrants in $H^{\bar{M}}$: $M_L P_{LH} \prod_{r=1}^{T^{\bar{M}}} P_{HHr} = P_{HL} M_H$. 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:

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

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

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.[40]

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 on 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.

Appendix C A model with undercapitalized newborn firms

This model is analogous to the baseline model, except for the following modifications. First, there is no productivity volatility, and all firms have the same productivity. Second, each period in each sector, qM_s firms die and are replaced by newborn firms. The ex-ante probability that

[39] Note that transition probabilities $P_{HH\tau}$ are allowed to be constant across τ.

[40] 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.

any firm dies is constant at q. However, related to the assumptions on transition probabilities in the baseline model, the death probability is not independent across firms. Specifically, I assume that each period, the same number of firms of each "type" die. Here, the definition of firm types is identical to the types of high-productivity firms in Appendix B, and the death probabilities are identical to the transition probabilities from high-productivity to low-productivity in that appendix.

The dead firms together transfer a fraction of their last period earnings to the newborn firms, such that these are born with capital levels $k_{s0} \equiv \zeta \frac{\phi_s P^F Q^F}{M_s}$, where K_s is aggregate capital in sector s and $0 < \zeta < 1$. Finally, firm-owner is has the following intertemporal preferences at time t:

$$U_{ist} = \sum_{v=t}^{\infty} (q\beta)^{v-t} c_{isv}$$

Where β is the discount factor, q is the ex-ante probability a firm dies in any given period and c_{ist} is firm-owner consumption. Otherwise, the model is exactly as the baseline model. The optimization problem and the solution to the steady state equilibrium are therefore highly similar, except that there are no productivity differences anymore. Now, the only variation in marginal costs is coming from differences in capital levels. Otherwise, reaction functions are still as in equation (22). The solution to the steady state equilibrium then implies:

Lemma 4. *In steady state, the joint distribution of capital and productivity within a sector is as follows:*

- *When firms are unconstrained τ periods after their birth , then $k_{is\tau} = k_s^*$*
- *When firms are constrained τ periods after their birth, then $k_{is\tau} = G_{s\tau} k_0$, with $G_{s\tau} = \Pi_{s=0}^{\tau-1}(\frac{\pi_{sv}}{P^F k_{sv}} + 1 - \delta)$*

The initial capital growth equation is then:

$$\frac{\pi_{s\tau}}{P^F G_{s\tau} k_{s0}} = (\mu_{s\tau} - (1-\alpha)) \frac{w}{P^F(1-\alpha)} \frac{l_{s\tau}}{G_{s\tau} k_{s0}},$$

from which one can derive analogous comparative statics results as before. In particular, as M_s increases, $G_{s\tau}$ will decline, which is the prediction that I take to the data.