Heterogeneous Markups, Growth and Endogenous Misallocation

Michael Peters*
Yale University

June 2019

Abstract

Markups vary systematically across firms and are a source of misallocation. This paper develops a tractable model of firm-dynamics where firms' market power is endogenous and the distribution of markups emerges as an equilibrium outcome. Monopoly power is the result of a process of forward-looking, risky accumulation: firms invest in productivity growth to increase markups in their existing products but get stochastically replaced by more efficient competitors. Creative destruction therefore has pro-competitive effects because faster churn gives firms less time to accumulate market power. At the aggregate level, this lowers the level and the dispersion of markups and hence reduces misallocation. In an application to firm-level data from Indonesia, the model predicts that (relative to the US) misallocation is more severe and firms are substantially smaller. To explain these patterns, the model suggests an important role for frictions for existing firms to enter new markets. Differences in entry costs for new firms are less important. Quantitatively, the static aggregate efficiency losses of misallocation due to markups are modest relative to cross-country differences in aggregate TFP.

*eMail: m.peters@yale.edu. I am grateful to Daron Acemoglu, Abhijit Banerjee and Rob Townsend for their invaluable guidance and to Ufuk Akcigit, Susanto Basu, Penny Goldberg, Chad Jones, Pete Klenow, Sam Kortum, Giuseppe Moscarini, Ezra Oberfield and Aleh Tsyvinski as well as Gianluca Violante and four anonymous referees, whose suggestions benefited the paper substantially. I also thank numerous seminar participants for their helpful suggestions.

1 Introduction

Firms' market power is an important determinant of static allocative efficiency. Similarly, the welfare consequences of particular policies like trade liberalizations or changes in the costs of entry directly depend on whether and how firms' markups change as a response. Standard macroeconomic models of firm behavior, however, routinely abstract from such considerations and treat market power as exogenous. In this paper I propose a parsimonious theory of firm-dynamics, where the distribution of markups emerges as an equilibrium outcome and is jointly determined with the rate of aggregate productivity growth.

My theory builds on firm–based models of Schumpeterian growth in the spirit of Klette and Kortum (2004). These models stress the importance of creative destruction whereby firms grow at the expense of other producers through the accumulation of products new to them. I embed this framework into a model of imperfect product markets, where firms compete a la Bertrand, engage in non-competitive pricing and charge variable markups. By endogenizing the extent of market power in a firm-based model of growth, this paper offers a unified perspective on how markups, misallocation, the process of firm-dynamics and aggregate growth are related.

At the heart of the economic mechanism is the idea that firms improve their productivity to accumulate market power. When a firm starts producing a particular product, markups are low as Bertrand competition forces the firm to charge a limit price. Over time, the firm spends resources to increase its productivity, pulls away from its competitors and raises the markup it optimally charges. Markups within products are therefore increasing as long as the product is produced by a given firm. Once the firm gets replaced by a new, more productive producer, markups get "reset" as Bertrand competition intensifies. It is the combination of firms engaging in markup-increasing quality improvements in their existing products and the process of markup-reducing product churning induced by creative destruction which shapes the stationary distribution of markups.

The model can be solved analytically and hence allows for a precise theoretical characterization of both the underlying determinants of market power and its macroeconomic consequences. First I show that the unique stationary distribution of markups is a pareto distribution, whose shape parameter is an endogenous statistic which I call the churning intensity. This statistic is simply the rate of creative destruction, i.e. the rate at which firms get replaced in their existing markets, relative to the speed at which firms increase their market power. If churning is intense, the cross-sectional distribution of markups is compressed because new firms replace existing producers quickly and keep monopoly power limited. If the extent of churning is limited, the distribution of markups has a fat tail because incumbent firms have ample time to accumulate market power. Secondly, I show that this endogenous pareto tail fully determines the macroeconomic implications of market power. More specifically, both aggregate TFP and the labor share of production workers (which in my model are the two sufficient statistics for the aggregate effects of non-competitive output markets) can be written explicitly as functions of the pareto tail. These expressions make precise why creative destruction has pro-competitive effects: by increasing the churning intensity, it reduces misallocation, raises aggregate TFP and increases the labor share.

Despite its parsimony, the model has rich testable implications for the empirical patterns of markups, size and age at the firm-level. The model predicts markups to follow a distinct life cycle pattern within products: conditional on survival, markups increase stochastically as the result of the firm's productivity-enhancing investments. To aggregate these product-level dynamics to the firm-level, two counteracting effects are at play. On the one hand, firms increase markups in their existing products as they age. This "own-innovation channel" raises the average markup of old firms relative to young firms. On the other hand, firms also expand into new products and lose existing products to other firms. This "creative-destruction channel" lowers the average markup of old firms as markups in new products are on average lower than in the old, infra-marginal products the firm loses. The "own-innovation channel" therefore increases the growth of markups by age and reduces the extent to which firms grow in

size as they get older. The "creative-destruction channel" has the exact opposite effect: it increases the rate of life cycle employment growth at the expense of profitability. I derive an explicit formula for the life cycle of markups at the firm-level and show that the first channel dominates for the vast majority of firms. Hence, as in the data, both average markups and firm size are predicted to increase in age.

I apply the theory to a particular setting which is motivated by the recent literature on misallocation. Following the work by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), this literature typically takes the extent of misallocation across firms as given, measures the dispersion in marginal products as firm-specific wedges and then quantifies its effect on cross-country differences in aggregate TFP. My model is consistent with this measurement approach, i.e. markups are isomorphic to the inferred firm-specific wedges (what is sometimes referred to as "TFPR"). Such wedges, however, are determined endogenously, have a precise structural interpretation and respond to policies or counterfactual changes in structural parameters. When I calibrate my model to firm-level panel data from Indonesia, I find that markups can plausibly account for 15% of the measured dispersion in TFPR. If markups were the only source of misallocation, this would reduce aggregate TFP by roughly 1%. Hence, the static efficiency losses from markups alone are unlikely to be the main culprit of aggregate TFP differences.

To better understand the underlying determinants of misallocation, firm size and growth, I then consider a specific counterfactual exercise. I start from the premise that a developing economy like Indonesia might suffer from frictions that hamper firms' ability to enter new product markets. Such frictions can be related to policies like size-requirements or lengthy approval processes for production licenses. They could also be technological in nature, whereby the costs of breaking into markets firms previously did not cater to, are higher in developing countries. I use the model to quantify the importance of such frictions for the equilibrium degree of misallocation, the firm-size distribution and the aggregate rate of growth.

The model highlights that such frictions come in two flavors. While expansion costs for existing firms make it costly for incumbents to break into new product markets, entry costs distort the incentives of entirely new firms to start producing. The model allows me to identify cross-country differences in these frictions from the entry rate and the rate of life cycle employment growth. For example, firms in the US grow faster then their Indonesian counterparts but the rates of entry are roughly similar. These two moments imply that both entry and expansion costs are lower in the US. However, while I estimate entry to be about 15% less costly in the US, the difference in expansion costs for existing firms is twice as a large and amounts to 30%. The reason why the model infers that expansion costs are more important than entry costs is simple: compared to firms in Indonesia, firms in the US grow faster conditional on survival. This makes the average firm larger. Lower entry costs have exactly the opposite effect as new entrants compete with old firms for customers, thus slowing down the extent of life cycle growth and reducing average firm size.

As a result of these lower frictions, creative destruction is much more potent in the US. This has implications for the distribution of firm size, misallocation and aggregate growth. Quantitatively, I find that the aggregate importance of small firms declines by 75% and that average firm size more than doubles. Moreover, the increase in churning reduces misallocation relative to Indonesia by roughly one-third, i.e. it increases allocative efficiency by 0.3%. The implications for the growth rate are subtle. Higher entry and expansion costs in Indonesia reduce creative destruction, which lowers the equilibrium growth rate. At the same time, by increasing the survival probabilities for existing firms, such barriers raise the incentives for firms to increase productivity within their markets. This tends to increase the equilibrium growth rate, albeit at the cost of higher markups. In my calibration these two effects essentially cancel out. Cross-country differences in firm size and misallocation therefore do not necessarily imply that the distribution of income across countries diverges in the long-run.

Related Literature The theory presented in this paper is an endogenous growth model in the Schumpeterian tradition of Aghion and Howitt (1992) and Grossman and Helpman (1991). In terms of modeling choices I build heavily on Klette and Kortum (2004). This framework is analytically attractive, can rationalize many salient features of the data (Lentz and Mortensen, 2008; Akcigit and Kerr, 2018) and has been used to study industrial policies (Acemoglu et al., 2016; Atkeson and Burstein, 2015), to quantify the importance of managerial delegation (Akcigit et al., 2015), to analyze the optimal protection of intellectual property rights (Acemoglu and Akcigit, 2012) and to measure the sources of US growth (Garcia-Macia et al., 2016). I show how to extend this framework in a tractable way to endogenize the distribution of markups. This extra margin does not only generate additional testable predictions but also has novel aggregate implications as the extent of misallocation and the aggregate rate of growth are jointly determined.

A growing empirical literature shows that markups vary systematically across firms and respond to changes in the environment. In particular, markups are low for entering firms (Foster et al., 2008), high for exporters (De Loecker and Warzynski, 2012) and increase in response to trade liberalizations (De Loecker et al., 2016). Recently, there has also been a growing interest in the macroeconomic implications of market power. De Loecker and Eeckhout (2017) and Autor et al. (2017) for example argue that markups have been rising and that highly profitable "superstar" firms have become more important for the aggregate economy. A recent literature also analyzes the consequences of the declining dynamism in the US on markups, the aggregate labor share and misallocation (Decker et al., 2014; Haltiwanger et al., 2015; Akcigit and Ates, 2019; Aghion et al., 2019; Peters and Walsh, 2019; Edmond et al., 2018). The theory proposed in this paper is qualitatively consistent with these trends as it predicts that profitability rises over the firms' life-cycle and and that markups, concentration and misallocation increase as a response to a decline in creative destruction and churning.

The application in this paper is motivated by the literature on misallocation (see e.g. Hsieh and Klenow (2009), Restuccia and Rogerson (2008), Bartelsman et al. (2013), David and Venkateswaran (2019) and the survey article by Hopenhayn (2012)). This literature typically treats misallocation as an exogenous firm-specific wedge. The model proposed in this paper generates misallocation endogenously. While numerous theories of misallocation based on inefficient input use have been proposed (e.g. imperfect capital markets (Buera et al., 2011; Moll, 2014; Midrigan and Xu, 2014), information frictions (David et al., 2016) or adjustment costs (Asker et al., 2014)), this paper is the first to structurally explore the role of monopolistic power as a source of misallocation. A recent literature has also studied how exogenous distortions affect firm growth and entry (see e.g. Hsieh and Klenow (2014), Bento and Restuccia (2017), Da-Rocha et al. (2017), Fattal Jaef (2018) and Buera and Fattal Jaef (2016)). In this paper, I take the opposite approach as misallocation is fully endogenous and hence emerges as an equilibrium outcome together with the size distribution of firms.

Finally, there is a growing literature in the field of international trade stressing the importance of markups. On the theory side, Bernard et al. (2003), Melitz and Ottaviano (2008) or Atkeson and Burstein (2008) are examples of theories that generate heterogeneous markups. That the welfare gains from trade are affected by markup heterogeneity and misallocation is explicitly analyzed in Edmond et al. (2015), Epifani and Gancia (2011), Arkolakis et al. (2016), De Blas and Russ (2015) and Holmes et al. (2014). In contrast to the model of this paper, all these frameworks assume that firm efficiency is exogenous.

The rest of the paper proceeds as follows. In the next section I present the theory and show how the joint distribution of markups and firm size is determined in equilibrium. Section 3 applies the theory to Indonesian micro data and quantifies the macroeconomic effects of markup heterogeneity. Section 4 concludes. The Appendix contains most proofs of the theoretical results and additional details of the empirical analysis.

2 Theory

2.1 The Environment

There is a measure one of infinitely lived households that supply their unit time endowment inelastically. Individuals have preferences over the unique consumption good, which are given by

$$U = \int_{t=0}^{\infty} e^{-\rho t} \ln(c_t)\, dt.$$

This final good, which I take to be the numeraire, is a Cobb-Douglas composite of a continuum of differentiated products

$$\ln Y_t = \int_0^1 \ln\left(\sum_{f \epsilon S_{it}} y_{fit}\right) di,$$

where y_{fit} is the quantity of product i bought from firm f and S_{it} denotes the number of firms competing in market i at time t. Hence, different products i and i' are imperfect substitutes, whereas there is perfect substitutability between different firms within a product. The assumption of a unitary demand elasticity is convenient for tractability. In Section 2.9 I show how the analysis can be extended to a more general setting.

Firms can produce multiple products and the only source of heterogeneity across firms is their factor-neutral productivity to produce different products. In particular, a firm f producing product i with productivity q_{fi} produces output according to

$$y_{fi} = q_{fi} l,$$

where l is the amount of labor hired. The market for intermediate goods is monopolistically competitive, so that firms take aggregate prices as given. In contrast, firms compete a la Bertrand with producers offering the same variety. This strategic interaction across producers is the source of heterogeneous markups and aggregate misallocation.

Both the set of competing firms $[S_{it}]_i$ and firms' productivities $[q_{fit}]_{fi}$ evolve endogenously through (i) the entry of new producers into the economy, (ii) the expansion of existing firms into new markets, i.e. into varieties they did not produce before and (iii) productivity increases by current producers in markets they already serve. While the first two margins of growth are considered in Klette and Kortum (2004), the third aspect is novel. It is this intensive margin of own-innovation that allows firms to gain competitiveness relative to other firms and gives rise to heterogeneous markups across producers. At the aggregate level, this ingredient provides the link between growth, misallocation and the process of firm-dynamics.

2.2 Static Allocations: Markups and Misallocation

Consider first the static allocations taking the number of firms and distributions of productivity q as given. To simplify the notation I will drop the time subscripts. Because production takes place with a constant returns to scale technology, firms compete in prices and different brands of product i are perceived as perfect substitutes, in equilibrium only the most productive firm within a product market will be active. However, the presence of competing producers (even though they are less efficient) imposes a constraint on the leading firm's price setting and forces the most efficient firm to resort to limit pricing. Letting q_i denote the productivity of the actual producer,

i.e. the most efficient firm within the market, the equilibrium markup for product i is given by

$$\mu_i \equiv \frac{p_i}{w/q_i} = \frac{w/q_i^F}{w/q_i} = \frac{q_i}{q_i^F},$$

where w denotes the equilibrium wage and w/q_i^F is the marginal cost of the second most productive firm, which I refer to as the follower. Hence, the equilibrium markup is simply given be the productivity advantage over competing firms.[1] To derive employment and profits at the product level, note that the Cobb-Douglas assumption on consumers' preferences implies that sales are equalized across markets, i.e. $p_{it}y_{ift} = Y_t$, so that

$$l_i = \frac{1}{q_i}y_i = \frac{1}{q_i}\frac{Y}{p_i} = \mu_i^{-1}\frac{Y}{w} \quad \text{and} \quad \pi_i = \left(1-\mu_i^{-1}\right)Y, \tag{1}$$

i.e. both employment l_i and profits π_i only depend on the markup μ_i and not on the level of productivity q_i. As I show below, this property makes the characterization of the model extremely tractable.

To derive the allocation of labor at the firm-level, let N_f be the set of products firm f produces. Total employment of firm f, l_f, is then given by

$$l_f = \sum_{i\in N_f} l_i = \sum_{i\in N_f} \mu_i^{-1}\frac{Y}{w} = \frac{Y}{w}n_f\mu_f^{-1} \quad \text{where} \quad \mu_f = \left(\frac{1}{n_f}\sum_{i\in N_f}\mu_i^{-1}\right)^{-1}, \tag{2}$$

where $n_f = |N_f|$ is the number of products produced by firm f and μ_f is the markup at the firm-level, which is simply the geometric average of firm f's product level markups. Equation (2) highlights that the size of a firm is shaped by two forces. Firms are large if they produce many products, i.e. expanding into novel markets increases employment at the firm-level. Conversely, higher markups reduce firm-employment holding the number of markets fixed.

Given the above structure, the economy has a transparent aggregate representation. Letting L_{Pt} denote the total mass of production workers, (2) implies that

$$L_P = \int_f l_f df = \frac{Y}{w}\int_f \sum_{i\in N_f}\mu_i^{-1}df = \frac{Y}{w}\times\left(\int_0^1 \mu_i^{-1}di\right).$$

Similarly, given that the final good is the numeraire, equilibrium wages are

$$w = \exp\left(\int_0^1 \ln q_i^F di\right) = \exp\left(\int_0^1 \ln\frac{q_i}{\mu_i}di\right) = Q\times\exp\left(\int_0^1 \ln\mu_i^{-1}di\right),$$

where $\ln Q = \int_0^1 \ln q_i di$ is the usual CES efficiency index. Aggregate output is therefore given by

$$Y \;=\; Q\mathcal{M}L_P \quad \text{where} \quad \mathcal{M} \equiv \frac{\exp\left(\int_0^1 \ln\mu_i^{-1}di\right)}{\int_0^1 \mu_i^{-1}di} = \frac{\exp\left(E\left[\ln\mu_i^{-1}\right]\right)}{E\left[\mu_i^{-1}\right]}. \tag{3}$$

Equation (3) highlights the macroeconomic consequences of market power. Aggregate TFP is the product of the physical productivity measure Q and the term $\mathcal{M}$, which summarizes the degree of markup misallocation. In the absence of markups, i.e. $\mu_i = 1$, it follows that $\mathcal{M} = 1$. Moreover, it is easy to show that $\mathcal{M} \leq 1$ and that $\mathcal{M} = 1$ if

[1] That markups are fully determined from limit pricing makes the dynamic decision problem of the firm very tractable. In Section OA-1.3 in the Online Appendix I discuss more specifically why the model would be far less tractable if firms were assumed to compete a la Cournot instead of Bertrand.

and only if markups are equalized. Hence, aggregate TFP depends on the *dispersion* of markups. While a common proportional increase in markups leaves the degree of misallocation unchanged, higher markup dispersion reduces allocative efficiency and hence aggregate TFP.

Monopoly power does not only affect aggregate TFP but also factor prices through a reduction in labor demand. In particular, equilibrium wages are distorted relative to their social marginal product and satisfy

$$\Lambda \equiv \frac{wL_P}{Y} = \left(\int_0^1 \mu_i^{-1} di\right) = E\left[\mu_i^{-1}\right]. \tag{4}$$

In contrast to $\mathcal{M}$, Λ depends on the level of markups, i.e. Λ is homogeneous of degree minus one in firms' markups μ_i. Note also that the canonical case of constant markups as generated by a CES demand system with differentiated products is a special case of these result: TFP is identical to its competitive counterpart but monopolistic power reduces factor prices.

Equations (3) and (4) highlight that the static macroeconomic implications of market power are fully summarized by two aggregate statistics $\mathcal{M}_t$ and Λ_t. Moreover, both statistics only depend on the marginal distribution of markups.[2] Below I construct this marginal distribution as an endogenous outcome from firms' innovation decisions so that misallocation, factor shares and growth are jointly determined in equilibrium.

Markups, Misallocation, Wedges and Measured Productivity The theory makes precise predictions about the link between physical productivity, measured productivity and the allocative consequences of markup power. This is reminiscent of the recent literature on misallocation pioneered by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), which stresses that resources are misallocated if revenue productivity ("TFPR") varies across producers. In my model, the variation in revenue productivity across producers is fully determined by the variation in markups, because revenue productivity in product i is given by

$$TFPR_{it} \equiv \frac{p_{it}y_{it}}{w_{rt}l_{it}} = \frac{p_{it}q_{it}}{w_{rt}} = \mu_{it}. \tag{5}$$

Hence, revenue productivity is equalized if and only if producers post a common markup. In the frameworks of Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), revenue productivity is determined by exogenous distortions ("firm-specific wedges"). Equation (5) shows that a firm charging a high markup has high revenue productivity and would be identified as facing a high distortionary tax. Intuitively: while market power makes firms too small, they might not be constrained or subject to policy frictions but rather choose to under-produce by setting a high price.

Equation (5) also highlights two aspects of my theory, which are absent from the above mentioned literature on misallocation. Most importantly, the markup μ_{it} is not an exogenous fundamental but is endogenously determined with the evolution of firms physical efficiencies. Secondly, the aggregate statistics $\mathcal{M}$ and Λ depend on the distribution of markup across *products*. Empirically, revenue productivity is usually measured at the *firm*-level. The mapping between the macroeconomic consequences of misallocation and the firm-level data therefore depends crucially on the distribution of firm size. In particular, in environments where firms produce multiple products, the measured dispersion in *firm*-level markups will underestimate the dispersion of markups across products, which is welfare-relevant.

[2] This is a consequence of the Cobb-Douglas structure. In Section 2.9 I generalize the analysis to the case of CES preferences and show that in that case the joint distribution of firm productivity q_i and markups μ_i is required.

2.3 Dynamics: Innovation and Creative Destruction

Both the production possibility frontier Q_t and the distribution of markups depend on the underlying distribution of productivity across firms. Following Aghion and Howitt (1992), Grossman and Helpman (1991) and Klette and Kortum (2004), I model firms' efficiencies as being ordered on a quality ladder with proportional productivity improvements of size $\lambda > 1$. Specifically, letting r denote the rung of the ladder, qualities are ordered according to $q_{r+1} = \lambda q_r$.This structure is convenient because it implies that equilibrium markups are given by

$$\mu_{it} = \frac{q_{it}}{q_{it}^F} = \frac{\lambda^{r_{it}}}{\lambda^{r_{it}^F}} = \lambda^{r_{it} - r_{it}^F} \equiv \lambda^{\Delta_{it}}, \tag{6}$$

where r_{it} and r_{it}^F denote the respective rungs on the quality ladder and $\Delta_{it} = r_{ti} - r_{it}^F \geq 1$ summarizes the producer's productivity advantage in market i. Hence, there is a one-to-one mapping between the quality gap Δ and the equilibrium markup μ.

The current level of productivity of product i, q_{it}, can increase in three ways: (i) a new firm can enter product i with a superior technology ("creative destruction by entrants"), (ii) an existing firm, who is not currently active in market i, can expand into this market ("creative destruction by incumbents") and (iii) the current producer of product i can increase its productivity to gain additional monopoly power ("own-innovation"). I assume that these three sources of growth are symmetric in that they improve the current frontier productivity by a single step from q_{it} to λq_{it}. This assumption is not only standard in most Schumpeterian models of growth but is particularly appealing in the current context as it highlights the different allocative consequences of creative destruction and own-innovation. In case the innovation stems from the current producer of product i the equilibrium markup *increases* by a factor λ as the quality gap rises from Δ to $\Delta + 1$. In contrast, when productivity growth is due to creative destruction, the equilibrium markup *decreases* by a factor $\lambda^{\Delta(i,t)-1}$, as the new producer is only a single step ahead on the quality ladder so that the quality gap declines from Δ to unity.[3]

To characterize the equilibrium rates of own-innovation, incumbent creative destruction and entry, I of course need to solve for the value function as all these choices are forward-looking. In principle, the state of a firm is consists of the quality its products $[q_j]_{j=1}^n$ and the quality-gaps in each product line $[\Delta_j]_{j=1}^n$. However, equations (1) and (6) imply that equilibrium profits of product i are given by $\pi_{it} = \left(1 - \lambda^{-\Delta_{it}}\right) Y_t$, i.e. they only depend on the quality gap. I therefore restrict attention to equilibria where firm behavior only depends on the payoff-relevant state variables $\left(n, [\Delta_j]_{j=1}^n\right)$. I adopt the usual stochastic formulation whereby firms can chose the flow rates of increasing the productivity of existing products and of expanding in a novel, randomly selected market at a cost. I denote the rate of own-innovation on existing products by $[I_{it}]_{i=1}^n$ and the rate of expansion by $[x_{it}]_{i=1}^n$. The associated cost function (denoted in units of labor) is given by $\Gamma\left([x_i, I_i]_{i=1}^n; n, [\Delta_j]_{j=1}^n\right)$.[4] Optimal behavior is then described by the value function $V_t\left(n, [\Delta_i]\right)$, which is given by

[3] In Section 2.9 I generalize the analysis to a setting where the step size is not necessarily equal to unity but drawn from a distribution. Note also that the continuous time formulation of the model precludes the possibility that a variety experiences both entry and a productivity improvement by the current producer.

[4] I formulate the firm's problem in terms of the expansion rates *per product*, x_i. This is for notational simplicity only. In equilibrium, the per product expansion rates are going to equalized. Hence, an equivalent formulation assumes that the firm as a whole choses an expansion rate $X = \sum x_i = nx$.

$$
\begin{aligned}
r_t V_t\left(n,[\Delta_i]\right)-\dot{V}_t\left(n,[\Delta_i]\right) \quad = \quad & \underbrace{\sum_{i=1}^{n}\pi_t\left(\Delta_i\right)}_{\text{Flow profits}}-\underbrace{\sum_{i=1}^{n}\tau_t\left[V_t\left(n,[\Delta_i]\right)-V_t\left(n-1,[\Delta_i]_{j\neq i}\right)\right]}_{\text{Creative destruction}}+ \\
& \max_{[x_i,I_i]_{i=1}^{n}}\left\{\underbrace{\sum_{i=1}^{n}I_i\left[V_t\left(n,\left[[\Delta_i]_{j\neq i},\Delta_i+1\right]\right)-V_t\left(n,[\Delta_i]\right)\right]}_{\text{Own-innovation}}+\right. \\
& \left.\underbrace{\sum_{i=1}^{n}x_i\left[V_t\left(n+1,[[\Delta_i],1]\right)-V_t\left(n,[\Delta_i]\right)\right]}_{\text{Expansion in new markets}}-\underbrace{\Gamma\left([x_i],[I_i];n,[\Delta_i]\right)w_t}_{\text{Innovation costs}}\right\}.
\end{aligned}
$$

The value of the firm consists of three parts. First there is the total flow payoff, which is simply the sum of profits across all markets. Second, there is the possibility of losing any of the n existing products to other firms. This happens at the endogenous rate of creative destruction τ, which is determined in equilibrium. Finally, there is an option value terms, which captures the possibility to increase the markup in each of its products through own-innovation and to expand into new markets. Note that the quality gap in novel markets is always equal to unity, i.e. new products have low markups. The last term captures the cost of own-innovation and market expansion.

While this expression for V_t looks daunting, it turns out that V_t admits a simple closed form solution. I assume a particular, additively separable functional form for the cost function $\Gamma(.)$, which is consistent with balanced growth and allows me to derive an analytic solution

$$
\Gamma\left([x_i],[I_i];n,[\Delta_i]\right) \quad = \quad \sum_{i=1}^{n}c\left(I_i,x_i;\Delta_i\right) \text{ where } c\left(I,x;\Delta\right)=\lambda^{-\Delta}\frac{1}{\varphi_I}I^{\zeta}+\frac{1}{\varphi_x}x^{\zeta}. \tag{7}
$$

Here φ_I and φ_x parametrize the efficiency of the innovation and expansion technology and $\zeta>1$ ensures that the cost function is convex so that there is a unique solution. The cost shifters φ_I and φ_x comprise both technological features of the innovation process as well as institutional determinants like bureaucratic requirements to produce a particular product. In addition, both cost functions contain scaling variables, which make the model consistent with balanced growth (Sutton, 1997; Luttmer, 2010).[5]

As far as new entrants are concerned I assume that potential entrants have access to a linear entry technology, whereby each unit of labor generates a flow of φ_z marketable ideas. As firms enter in a single market with a unitary quality gap, the free entry condition is given by

$$
V_t\left(1,1\right)\leq\frac{1}{\varphi_z}w_t=0 \text{ with equality if } z>0,
$$

where z denotes the equilibrium flow rate of entry. For the remainder of the paper, I focus on the case with positive entry where the free entry condition holds with equality. The aggregate rate of creative destruction τ_t is therefore

[5]The term $\lambda^{-\Delta}$ in $c^I(.)$ implies that innovations are easier the bigger the within-market productivity advantage Δ. This is similar in spirit to the assumption of knowledge capital made in Klette and Kortum (2004) or the setup in Atkeson and Burstein (2010). Intuitively: per-period profits are given by $\left(1-\lambda^{-\Delta}\right)Y$ and hence concave in Δ. For innovation incentives to be constant, the marginal costs of innovation have to be lower for more advanced firms. The leading term in (7) is exactly the right normalization to balance those effects. Note that firms only generate a high productivity gap when they have multiple innovation *in a row*. Hence, (7) effectively posits that firms can build on their own innovations of the past. The term $n^{1-\zeta}$ in $c^X(.)$ serves a similar purpose and implies that the cost of expanding at rate x per market (i.e. $nx=X$) is linear in n.

given by

$$\tau_t = z_t + \int_{i=0}^{1} x_{it} di,$$

as the producer of an individual product can be replaced both by entering firms and through the expansion of existing firms.

2.4 The Stationary Equilibrium

Given this set-up, I now characterize the stationary (or balanced-growth-path) equilibrium of this economy. Labor market clearing requires that production labor L_{Pt} and research labor L_{Rt} add up to the aggregate labor endowment, which I normalize to unity:

$$1 = L_{Pt} + L_{Rt} = L_{Pt} + \frac{1}{\varphi_z} z_t + \int_{i=0}^{1} \left(\frac{1}{\varphi_x} x_{it}^{\zeta} + \lambda^{-\Delta_i} \frac{1}{\varphi_I} I_{it}^{\zeta} \right) di.$$

A stationary equilibrium is then defined in the usual way.

Definition. A stationary equilibrium is a set of allocations $[l_{it}, I_{it}, x_{it}, z_t, y_{it}, c_t]_{it}$ and prices $[w_t, r_t, p_{it}]_{it}$ such that (i) all aggregate variables grow at a constant rate, (ii) consumers chose $[y_{it}, c_t]_{it}$ to maximize utility, (iii) firms chose $[I_{it}, x_{it}, p_{it}]$ optimally, (iv) the free entry condition is satisfied, (v) all markets clear and (vi) the cross-sectional distributions of markups and firm size are stationary.

Despite the fact that firms' market power is endogenous and that the model allows for both own-innovation and creative destruction by existing firms, the model is very tractable and the stationary equilibrium can be characterized analytically. Two properties are important for this result. First of all, the theory admits closed form solutions for the value function V_t and incumbents' innovation behavior. This is the content of Proposition 1. Secondly, the distribution of markups, which is endogenous and required to calculate the misallocation wedge $\mathcal{M}$ and the labor wedge Λ, can also be characterized analytically and I will do so in Section 2.5.

Proposition 1. *Consider the setup described above. Suppose that* $\rho > \frac{\zeta-1}{\zeta} \left(\frac{1}{\zeta} \frac{\varphi_x}{\varphi_z} \right)^{1/(\zeta-1)}$. *Then there exists a unique stationary equilibrium, where:*

1. *The value function is given by*

$$V_t\left(n, [\Delta_i]_{i=1}^{n}\right) = \sum_{i=1}^{n} V_t\left(\Delta_i\right) = V_t^P n + \sum_{i=1}^{n} V_t^M\left(\Delta_i\right) \tag{8}$$

where

$$V_t^P = \frac{\pi(1) + (\zeta - 1) \frac{x^{\zeta}}{\varphi_x} w_t}{\rho + \tau} \quad \text{and} \quad V_t^M(\Delta) = \frac{\pi(\Delta_i) - \pi(1) + (\zeta - 1) \lambda^{-\Delta_i} \frac{I^{\zeta}}{\varphi_I} w_t}{\rho + \tau},$$

2. *The optimal rates of innovation, expansion, entry and creative destruction,* $([I_{it}, x_{it}]_i, z_t, \tau_t)$, *are constant and given by* (I, x, z, τ). *In particular, the optimal expansion rate* x *is given by*

$$x = \left(\frac{\varphi_x}{\varphi_z} \frac{1}{\zeta} \right)^{\frac{1}{\zeta-1}}, \tag{9}$$

and the innovation rate I *solves*

$$I = \left(\frac{\lambda - 1}{\lambda} \frac{1}{\rho + \tau} \left(\frac{\varphi_I}{\zeta} \frac{Y_t}{w_t} - \frac{\zeta - 1}{\zeta} I^{\zeta} \right) \right)^{\frac{1}{\zeta-1}}, \tag{10}$$

3. The distribution of markups is stationary so that the equilibrium wedges $\mathcal{M}$ and Λ are constant,

4. All aggregate variables grow at the common growth rate

$$g = \frac{\dot{Q}_t}{Q_t} = \ln \lambda \times (I + x + z) = \ln \lambda \times (I + \tau).$$

Proof. See Section A-1.1 in the Appendix. The condition $\rho > \frac{\zeta-1}{\zeta}\left(\frac{1}{\zeta}\frac{\varphi_x}{\varphi_z}\right)^{1/(\zeta-1)}$ is sufficient for the free entry condition to be satisfied. □

Proposition 1 establishes that the economy permits a unique stationary equilibrium, which can essentially be characterized analytically. The value of the firm, $V_t(n, [\Delta_i])$, has an intuitive structure. First of all, it is additive across products. Secondly, the value of producing a given product with quality gap Δ_i also consists of two additive parts. The first term, V_t^P, captures the value of producing a particular product with a quality gap of unity (and hence a markup of λ). It consists of the production value and the inframarginal rents of the concave expansion technology, i.e. the option value of being able to expand in new markets. This part of a firm's value scales linearly in the number of markets n and is similar to the baseline model of Klette and Kortum (2004). The second term $V_t^M(\Delta)$ is novel and captures the value of the option to accumulate market-power. It consists of the flow value of being able to charge higher markups $(\pi(\Delta_i) - \pi(1))$, augmented by the possibility of increasing markups even further in the future. Because, firms are long-lived, the value function is given by the net-present value of these payoffs, where the appropriate discount rate is not only the rate of time preference ρ, but also the rate of creative destruction τ to account for the risk of being replaced.

Associated with the value function V_t are optimal innovation, entry and expansion decisions, which are constant - both across products and across time. Equation (9) shows that the rate at which incumbents enter new markets has a simple closed form expression and depends on the efficiency of incumbent creative destruction (φ_x) relative to the one of entrants (φ_z). The optimal extent of own-innovation I depends on two endogenous aggregate variables - the rate of creative destruction τ and size of the market relative to the cost of innovation $\frac{Y_t}{w_t}$. An increase in the rate of creative destruction τ reduces firms' incentives to engage in own-innovation, as the expected time-horizon of the accrual of monopolistic rents becomes shorter. Conversely, innovation incentives are high if aggregate demand Y_t is large relative to the cost of innovation w_t.

The equilibrium entry rate z is then determined from the labor market clearing condition and the rate of creative destruction is given by $\tau = z + x$.[6] The aggregate growth rate is simply given by the growth rate of Q_t, because the efficiency wedge $\mathcal{M}_t$ is constant in a stationary equilibrium. Because all three sources of innovation generate productivity improvements of the size λ, g is proportional to the sum of creative destruction τ and firms' rate of own-innovation I.

2.5 The Cross-Sectional Distributions of Markups

As highlighted in Proposition 1, the distribution of markups and hence the degree of misallocation summarizes by $\mathcal{M}$ and Λ is endogenous and determined the equilibrium. To construct this distribution, recall that markups only depend on the distribution of quality gaps Δ across products. The cross-sectional distribution of markups is therefore fully characterized by $\{\nu(\Delta, t)\}_{\Delta=1}^{\infty}$, where $\nu(\Delta, t)$ denotes the measure of products with quality gap Δ

[6] More specifically, together with the free entry condition and the equilibrium labor wedge Λ_t in (4), the labor market clearing condition and the three relationships (8), (9) and (10) are six equations in the six unknowns $\left(x, \tau, I, \frac{V_t}{w_t}, \frac{Y_t}{w_t}, L_{Pt}\right)$. In Section A-1.1 I show how these can be reduced to a system of two equations in two unknowns.

at time t. These measures solve the set of differential equations

$$\dot{\nu}(\Delta, t) = \begin{cases} -(\tau + I)\,\nu(\Delta, t) + I\nu(\Delta - 1, t) & \text{if } \Delta \geq 2 \\ \tau(1 - \nu(1, t)) - I\nu(1, t) & \text{if } \Delta = 1 \end{cases}, \tag{11}$$

where $\dot{\nu}(\Delta, t)$ denotes the time derivative. Intuitively, there are two ways for product i to leave state (Δ, t): the current producer could have an innovation (in which case the quality gap would increase from Δ to $\Delta + 1$) or a new producer could enter (in which case the quality gap would decrease to unity). The only way for a product to enter the state (Δ, t) is by being in state $\Delta - 1$ and then having the current producer experience an increase in productivity (which happens at rate I). The state $\Delta = 1$ is special, because all products where the producing firm gets replaced enter this state.

Equation (11) is the key equation to characterize the equilibrium distribution of markups. Three properties are noteworthy. First of all, the distribution is fully determined from the two endogenous variables (I, τ) and is hence jointly determined with the economy-wide growth rate g. Secondly, the distribution of firm size is not required to solve for the distribution of markups across products. This is due to the fact that all firms innovate and expand at constant rates I and x per market. Finally, (11) highlights the pro-competitive effects of creative destruction: while productivity growth by existing producers are markup increasing, market churning through creative destruction shifts the distribution of markups downwards. This suggests that creative destruction is a force that tends to reduce misallocation and that firms' own-innovation efforts lower allocative efficiency through higher and more dispersed markups. The next proposition shows that this intuition is exactly correct.

Proposition 2. *Let I and τ be the equilibrium rates of own-innovation and creative destruction in a stationary equilibrium. Let*

$$\theta = \frac{\ln(1 + \vartheta_I)}{\ln \lambda} \quad \textit{where} \quad \vartheta_I = \frac{\tau}{I}.$$

Then:

1. *The distribution of markups is given by*

$$G(\mu) = 1 - \mu^{-\theta},$$

2. *The aggregate misallocation measures $\mathcal{M}$ and Λ are given by*

$$\mathcal{M} = e^{-1/\theta}\frac{1 + \theta}{\theta} \quad \textit{and} \quad \Lambda = \frac{\theta}{1 + \theta}. \tag{12}$$

Proof. See Section A-1.2 in the Appendix. □

Proposition 2 contains the main theoretical result of this paper: the cross-sectional distribution of markups $G(\mu)$, the extent of misallocation $\mathcal{M}$ and the labor wedge Λ are jointly determined with firms' innovation incentives and the rate of creative destruction. In particular, the endogenous distribution of markups takes a pareto form, whose shape parameter θ is endogenous and fully determined from a single endogenous statistic - the *churning intensity* ϑ_I. This statistic measures the speed with which firms are being replaced by new producers *relative* to firms' own-innovation efforts. If churning is intense, the shape parameter is large so that both markup heterogeneity and the average markup decline. If in contrast churning is of little importance, the resulting distribution of markups has a fat tail and both the average markup and their dispersion is large. Because firms' own-innovation incentives I and the rate of creative destruction τ are determined endogenously and hence depend on parameters or policies,

markups and misallocation will also endogenous respond. This is the crucial difference to Bernard et al. (2003), who generate a Pareto distribution of markups from firms' exogenous productivity draws.[7]

The macroeconomic consequences of market power are fully summarized the two sufficient statistics $\mathcal{M}$ and Λ which also only depend on ϑ_I and have the closed-form expressions given in (12).[8] It is easy to verify that both $\mathcal{M}$ and Λ are increasing in ϑ_I. This captures the pro-competitive effect of creative destruction: by reducing equilibrium markups, creative destruction reduces misallocation and increases TFP and equilibrium factor prices holding the productivity frontier Q_t fixed. Note also that the standard deviation of log markups is given by θ^{-1} and is hence also decreasing in ϑ_I. Hence, as stressed in the literature on misallocation, the dispersion in log TFPR co-moves with aggregate TFP.

The mechanism which generates the endogenous pareto tail in my model is akin to the city-size dynamics of Gabaix (1999). Markups *within a product* have an intuitive life cycle interpretation: as long as the current producer does not get replaced, markups stochastically increase. Once a new producer breaks into the respective product market, markups are "reset" to λ and the process begins afresh. In fact, as shown in Section A-1.3 the Appendix, the conditional distribution of quality gaps Δ as a function of the time a product is produced by a particular firm, which I refer to as "product age" a_P, is a Poisson distribution with parameter Ia_P

$$h_{\Delta+1}(a_P) = \frac{1}{\Delta!}(Ia_P)^{\Delta} e^{-Ia_P}. \tag{13}$$

This implies that the average log markup of a product conditional on being served by the same firm for a_P years is given by

$$E[\ln \mu | \text{product age}{=}a_P] = \ln \lambda (1 + Ia_P), \tag{14}$$

i.e. is increasing in age at a rate proportional to I. Hence, conditional on not being replaced, the distribution of markups continuously shifts outwards as incumbent firms engage in productivity improvements to rack up their monopoly power. This process of accumulation is faster the higher I.

Products, however, are not produced by the same firm for eternity. In particular, creative destruction limits how long existing firms can survive. Because producers of a given product are replaced at rate τ, the probability of producing a product for at least a_P years is given by $e^{-\tau a_P}$. Hence, the extent to which firms can accumulate market power depends crucially on the degree of creative destruction τ. If τ is high, it is rare to see firms serving a particular product market for a long time. The long-run distribution of markups is shaped by the interplay of these two processes, which lead to a pareto distribution.[9]

2.6 Markup Dynamics at the Firm-Level

Equations (13) and (14) characterize the life cycle dynamics of markups at the product level. These implications cannot be taken directly to the data because firms are a collection of many products. This presence of multi-product

[7] De Blas and Russ (2015) analyze a version of Bernard et al. (2003), where the distribution of markups responds to trade policy. They consider a static model but allow for a finite number of firms, who differ in their efficiency (drawn from an exogenous Fréchet distribution). They show that the distribution of markups depends on the number of competing firms and that a higher number of competitors decrease markups.

[8] Note that Δ is not a continuous variable but only takes integer values. For simplicity I treat markups as continuous when calculating M and Λ from $G(\mu)$. See Section A-1.2 in the Appendix for the closed form expressions for the discrete case.

[9] To see this intuitively, suppose that (14) were to hold deterministically, i.e. $\ln \mu = \ln \lambda + \ln \lambda \times Ia$. Then,

$$P[\mu > \mu_0] \quad = \quad P\left[a > \ln \frac{\mu_0}{\lambda} \times \frac{1}{\ln \lambda} \frac{1}{I}\right] = e^{\ln\left(\frac{\mu_0}{\lambda}\right)^{-\frac{1}{\ln \lambda}\frac{\tau}{I}}} = \left(\frac{\mu_0}{\lambda}\right)^{-\frac{1}{\ln \lambda}\frac{\tau}{I}},$$

which is a pareto distribution. Jones and Kim (2016) exploit a similar structure to argue that creative destruction limits income inequality by reducing the time entrepreneurs have to accumulate firm-specific human capital.

firms makes the evolution of firm-level markups subtle. Consider a firm of age a_f. On the one hand, old firms tend to have high markups as they are the only firms with the potential of having had enough time to build markups for a given product through a series of successful own-innovations. This "own-innovation channel" implies that markups and age should be positively correlated. On the other hand, old firms also had ample time to expand into new markets and lose products from their portfolio. And as markups in new, "marginal" products are lower than markups for the average product firms lose, this "creative destruction channel" tends to lower the extent to which markups increase in age. Hence, the model predict a life cycle pattern where firms accumulate market power in their existing products to increase profitability and add new, low markup products to their portfolio.

To see this more clearly, suppose that firms were to never horizontally expand (i.e. $x = 0$) and hence never serve more than a single market. In that case, only the own-innovation channel is at play and the age of the firm a_f directly corresponds to the age of the product a_P. Hence, the average log markup by *firm* age is also given by (14). Allowing firms to expand horizontally into new markets breaks this tight link between markups and firm age. It is nevertheless the case that one can still derive an analytical characterization of the markup dynamics at the firm level.

Proposition 3. *The average firm-level log markup* $\ln \mu_f$ *as a function of firm age* a_f *is given by*

$$E\left[\ln \mu_f | \textit{firm age} = a_f\right] = \ln \lambda \left(1 + I \times E\left[a_P | a_f\right]\right), \tag{15}$$

where

$$E\left[a_P | a_f\right] = \frac{1}{x}\left(\frac{\frac{1}{\tau}\left(1 - e^{-\tau a_f}\right)}{\frac{1}{x+\tau}\left(1 - e^{-(x+\tau)a_f}\right)} - 1\right)\left(1 - \phi\left(a_f\right)\right) + a_f \phi\left(a_f\right),$$

and

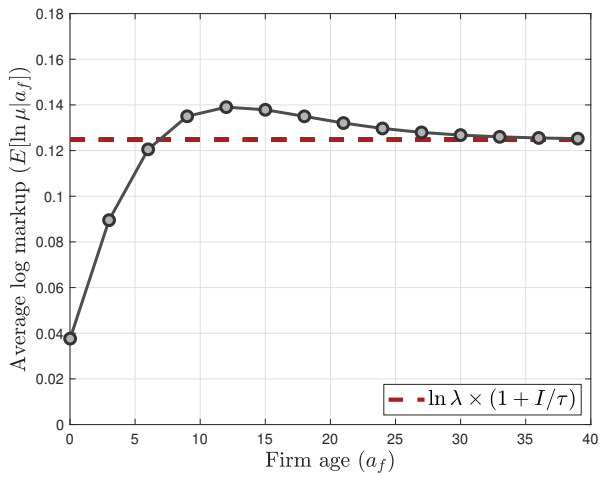
$$\phi\left(a\right) = e^{-xa}\frac{1}{\gamma\left(a\right)}\ln\left(\frac{1}{1-\gamma\left(a\right)}\right) \quad \textit{and} \quad \gamma\left(a\right) = \frac{x\left(1 - e^{-(\tau - x)a}\right)}{\tau - x \times e^{-(\tau - x)a}}.$$

Proof. See Section A-1.3 in the Appendix. □

Proposition 3 contains an analytic expression for the life cycle profile of markups. Note that, equation (15) has the same structure as (14), except that the mapping between firm age a_f and product age a_P is more complicated and depends on both the rate of incumbent expansion x and the extent of creative destruction τ. In particular, the possibility of firms breaking into new markets implies that $E\left[a_P | a_f\right] \leq a_f$. Moreover, it is easy to verify that $\lim_{x \to 0} E\left[a_P | a_f\right] = a_f$, so that (14) emerges as a special case.

The life cycle profile characterized in (15) is depicted in Figure 1. Surprisingly, the relationship is *non-monotone*. The intuition is the following. Recall that a given product is creatively destroyed with flow rate τ. Hence, the average survival time for a given product is τ^{-1}. For the set of very old firms, this is therefore exactly the average age of a given product in their portfolio, i.e. $\lim_{a_f \to \infty} E\left[a_P | a_f\right] = 1/\tau$. The limiting average markup for old firms (which is displayed in the red dashed line in Figure 1) is therefore given by $\ln \lambda \left(1 + I/\tau\right)$. Note that, as for the cross-sectional distribution of markups characterized in Proposition 2, the churning intensity $\vartheta_I = \tau / I$ again emerges as a crucial determinant for the level of markups. If the churning intensity ϑ_I is high, the average markup of old firms is low.

The reason why the average markup for younger firms deviates from this level is due to selection. For young firms, the age of the products they sell is obviously *negatively* selected - a two year old producer cannot possible sell a product that has been around for four years. And because markups increase in the average age of firms' product portfolios, young firms have low markups which are expected to increase. Interestingly, once firms become



Notes: The figure displays the expected log markup as a function of age, i.e. $E\left[\ln \mu_f | a_f\right]$ given in (15).

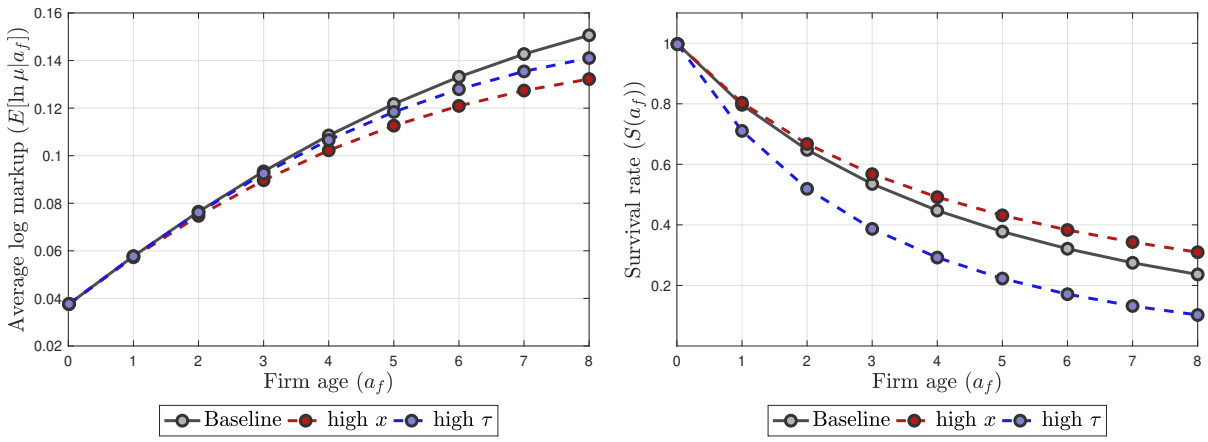
Figure 1: The Life Cycle of Markups

sufficiently old the expected age of the products they sell is *positively* selected. In particular, there is a chance that the firm still owns the product it initially started out with, which - for old firms - is older than the average product. In the limit this effect vanishes as the probability that a 40 year old firm managed to cling on to its initial product for 40 years goes to zero. Proposition 3 also highlights how the life cycle dynamics of markups are shaped by firms' endogenous innovation choices (I, x) and the rate of creative destruction τ. First of all, it is immediate that (for a given age) the average markup is increasing in I as $E\left[a_P | a_f\right]$ is independent of I, but a higher rate of own-innovation allows firms to increase their market power at a faster speed. The effects of incumbent expansion x and creative destruction τ are more subtle. In the left panel of Figure 2 I depict the effect of an increase in the expansion rate x (red line) and in the rate of creative destruction τ (blue line). Recall that $\tau = z + x$, i.e. holding x fixed an increase in the rate of entry translates one-to-one into higher creative destruction. For visual clarity I focus on the early part of a firms' life cycle, where markups are monotone in age. This is the empirically relevant case for my application.

While both τ and x *decrease* the extent of life cycle markup growth, the economics are very different. A higher rate of creative destruction reduces markup growth through higher churning so that firms have less time to accumulate market power in the products they own. In contrast, a higher expansion rate x reduces the average markup through a composition effect. If firms enter novel markets very frequently, only a small fraction of their sales is accounted for by old, high-markup products. This suggests that a higher expansion rate x reduces markups and increases allocative efficiency. However, as shown in Proposition 2, this is not the case as the distribution of markups is independent of x conditional on τ. To resolve this apparent contradiction, note that x and τ also affect firms' survival probabilities. Specifically, let $S\left(a_f\right)$ denote the share of firms surviving until age a_f. As I show in Section OA-1.2.3 in the Online Appendix,

$$S\left(a\right) \quad = \quad 1 - \frac{\tau}{x}\gamma\left(a\right) \tag{16}$$

where $\gamma\left(a\right)$ is given in Proposition 3. As seen in the right panel of Figure 2 (and shown formally in Section OA-1.2.3 in the Online Appendix), a higher rate of incumbent creative destruction x increases the number of surviving firms at every age bin. Hence, while a faster rate of market expansion x reduces the expected markup for a given age, it also increases the share of old firms, who have higher markups on average. In the cross-section, these two effects



Notes: The figure displays the expected log markup as a function of age (see (15)) in the left panel and the cumulative survival probability (see (16)) in the right panel. The baseline model is shown in grey. The red (blue) line refers to an increase in incumbent expansion x (creative destruction τ).

Figure 2: Life Cycle Dynamics of Markups and Survival

exactly cancel out, rendering the distribution of markups independent of x conditional on τ. In contrast, for the case of creative destruction, these two forces complement each other: not only is markup growth slower conditional on survival, but higher churning also implies that the age distribution shifts towards young, low-markup firms.

2.7 Sales and Employment Dynamics

The model also makes concise predictions for the dynamics of firm-size. First recall that aggregate sales at the firm level are proportional to the number of products n because of the Cobb-Douglas preferences. Second note that the stochastic process of firms losing and gaining products is the same as in Klette and Kortum (2004). This implies that the process of sales dynamics also takes exactly the same form. In particular the sales distribution is skewed, the variance of sales growth is decreasing in size, the probability of exit is declining in both size and age and Gibrat's Law will be a good approximation for the growth for large firms. Importantly, and in contrast to Klette and Kortum (2004), total sales and employment are no longer proportional as firm employment is also affected by the firm's average markup.[10] Interestingly, the determinants of the markup and cross-sectional sales distribution neatly separate. While the distribution of markups only depends on the churning intensity $\vartheta_I = \tau/I$, the distribution of sales can be fully characterized by the endogenous statistic $\vartheta_x = x/\tau$, i.e. the share of creative destruction, which is due to incumbent firms. In particular, as shown in Section A-1.4 in the Appendix, the number of active firms F is given by $F = \frac{1-\vartheta_x}{\vartheta_x} \ln\left(\frac{1}{1-\vartheta_x}\right)$ and the share of aggregate sales accounted for by firms with at most n products, Ω_n, is given by $\Omega_n = 1 - (\vartheta_x)^n$. This implies that if a large share of creative destruction is due to incumbents the number of active firms is small, average firm size is large and a large share of output is produced in large firms.

Because total employment depends both on firm sales and on the average markup, the dynamics of markups and firm size are informative about the relative importance of market expansion x and own-innovation I. In particular, as I show in Section A-1.4 in the Appendix, the average log employment by age $E[\ln l_f|a_f]$, i.e. the life cycle of

[10] Another difference is that my model predicts a deviation from the exact proportionality between R&D spending and sales. In particular, it is easy to verify that the R&D intensity, i.e. R&D spending as a fraction of sales, of firm f is proportional to $\varphi_x^{-1}x^{\zeta} + \varphi_I^{-1}I^{\zeta}\frac{1}{\mu_f}$, where μ_f is the firm-level markup defined in (2). Hence, small, young, low-markup firms tend to spend relatively more on R&D. This is qualitatively consistent with the findings reported in Akcigit and Kerr (2018). In absolute terms, R&D spending is of course increasing in size, i.e. the model generates systematic differences in R&D spending across firms.

firm size, is given by

$$E\left[\ln l_f|a_f\right] - E\left[\ln l_f|0\right] = \underbrace{\frac{1-\gamma\left(a\right)}{\gamma\left(a\right)}\sum_{j=1}^{\infty}\ln j \times \gamma\left(a\right)^j}_{\text{Sales growth}} - \underbrace{\ln\lambda\left(I \times E\left[a_P|a_f\right]\right)}_{\text{Markup growth}}, \tag{17}$$

where $\gamma\left(a\right)$ and $E\left[a_P|a_f\right]$ are characterized in Proposition 3. The more important firms' markup-increasing own-innovation I relative to their expansion rate x, the steeper the age-profile of markups and the flatter the extent of life cycle employment growth. If there is no scope for incumbent own-innovation (as in Klette and Kortum (2004)), sales and employment are proportional at all ages as markups are constant. If in contrast firms could not expand vertically they would only produce a single product and sales would not be a function of age. As markups increase in age, firm employment would decrease in age while markups and profitability increase. Hence, the relative speed at which firms increase their markups relative to their size identifies the relative importance of firms expanding their scope of production horizontally through creative destruction or vertically through productivity improvements in their existing products.

2.8 Determinants of Creative Destruction: Entry and Expansion Costs

The results above highlight that markups, productivity growth and the process of firm dynamics are jointly determined in equilibrium. Because the rate of creative destruction emerged as the key endogenous statistic, frictions for existing firms to expand into new product markets or costs for new firms to enter the economy are crucial determinants of misallocation, aggregate growth and the firm size distribution. In terms of the theory, such frictions are subsumed in the expansion and entry efficiency terms φ_x and φ_z. Consider for example the case of incumbent creative destruction x. The cost shifter φ_x parametrizes the resource requirements of coming up with the technology of producing a superior product and successfully introducing it to the market. Hence, a low level of φ_x can for example reflect license requirements or bureaucratic red tape, which have to be overcome before a firm can be active. Similarly, a low level of φ_z can capture lengthy approval processes for new producers to be allowed to produce. Because such frictions are thought to be particularly important in many developing countries, I study the manufacturing sector of Indonesia as an application of the theory.[11] The following proposition summarizes the effects of these frictions on the stationary equilibrium of the model.

Proposition 4. *Consider a stationary equilibrium and suppose that $\zeta \geq 2$. Higher entry and expansion costs reduce the churning intensity $\vartheta_I = \tau/I$ and creative destruction τ. Higher entry costs (expansions costs) increase (reduce) the incumbent creative destruction share $\vartheta_x = x/\tau$. Formally,*

$$\frac{\partial\vartheta_I}{\partial\varphi_z} > 0 \quad \frac{\partial\tau}{\partial\varphi_z} < 0 \quad \frac{\partial\vartheta_x}{\partial\varphi_z} > 0 \quad \textit{and} \quad \frac{\partial\vartheta_I}{\partial\varphi_x} > 0 \quad \frac{\partial\tau}{\partial\varphi_x} > 0 \quad \frac{\partial\vartheta_x}{\partial\varphi_x} > 0.$$

Higher entry and expansion costs therefore increase misallocation, i.e. reduce $\mathcal{M}$ and Λ. In contrast, entry costs (expansion costs) increase (reduce) average firm size F^{-1} and the output share of large firms Ω_n. The effect on the aggregate growth rate g is ambiguous.

Proof. See Section A-1.5 in the Appendix. The restriction that $\zeta \geq 2$ is a sufficient condition. □

[11] One widely used measure of entry costs is developed in Djankov et al. (2002). They measure the fees and time costs to legally operate a business for a variety of countries. Such variation in the regulation of entry has been linked to cross-country income differences in Barseghyan (2008), Barseghyan and DiCecio (2009) or Herrendorf and Teixeira (2011). There are also studies focusing on particular episodes of delicensing. The dismantling of India's Licence Raj, for example, has been studied in Aghion et al. (2008). Even though all these studies refer to "entry costs", the empirical variation is likely to capture both entry and expansion Costs in the sense of my theory. See in particular Bento (2016), who explicitly distinguishes between the cost of entry for new firms and existing firms.

Proposition 4 stresses that frictions for existing firms to enter new product markets and higher costs for new firms to enter the economy have qualitatively different implications. While both type of frictions reduce creative destruction and the churning intensity ϑ_I and hence increase misallocation, they affect the firm size distribution differentially. Expansion costs for existing firms reduce the share of creative destruction due to incumbents ϑ_x and lower average firm size. In contrast, higher entry costs increase ϑ_x, which leads to bigger but less firms being active in equilibrium.[12] This suggests that expansion costs might be of first-order importance in developing countries as firms in these economies are mostly small and misallocation is argued to be rampant. Entry costs in contrast face somewhat of an uphill battle, because their first-order effect on important moments like average firm size or the number of producers is counterfactual: high costs of entry firms larger in equilibrium.

It is also noteworthy to point out that the relationship between entry and expansion costs and the endogenous growth rate is ambiguous. If creative destruction and own-innovation are strong substitutes, it is possible that $I+\tau$, which - recall - determines the equilibrium growth rate, increases in response to higher entry or expansion costs. In fact, this substitutability is not only a theoretical possibility but turns out to be quantitatively important in the calibrated model.[13]

2.9 Theoretical Extensions

The baseline model laid out above can be solved analytically and highlights the economic mechanism in it cleanest form. This tractability of course required stringent assumptions. Of particular importance seem to be the restrictions of a constant step size, i.e. all innovations improve upon the frontier technology by a single step, and the unitary elasticity of demand embedded in the Cobb-Douglas structure of consumers' preferences. In this section I discuss in more detail to what extent these assumptions are consequential.

2.9.1 Stochastic Step Size

Consider first the assumption of a constant step size. It turns out that this assumption is not only easy to dispense with but that all my results directly apply to a more general environment. Suppose innovating firms improve upon the current producer by $\tilde{k}$ steps (each of size λ), where $\tilde{k}$ is a random variable with $p_j = P\left[\tilde{k} = j\right]$ and $\sum_{j=1}^{\infty} p_j = 1$. The baseline model is the special case with $p_1 = 1$. In Section A-1.6 in the Appendix I show that the model with this extension is as tractable as the baseline model. In particular, the value function can still be solved explicitly, the innovation and entry policies (I, x, z) are constant and the stationary equilibrium can be explicitly characterized. Crucially, the endogenous distribution of markups is very similar to the baseline model.

Proposition 5. *Let $\{p_j\}_j$ be the probability of increasing the frontier productivity by j steps conditional on innovating. Consider a BGP, where innovation, expansion and entry rates are constant and given by (I, x, z). Then:*

1. *The unique stationary distribution of quality gaps $\{\nu(\Delta)\}_{\Delta=1}^{\infty}$ is defined recursively as*

$$\nu_\Delta = \frac{1}{1+\vartheta_I}\left(\sum_{m=1}^{\Delta-1} \nu_m p_{\Delta-m}\right) + \frac{\vartheta_I}{1+\vartheta_I} p_\Delta \quad \text{for } \Delta = 1, 2, 3...$$

[12] As for the response of I there are two effects. On the one hand, the decline in creative destruction raises firms' incentives to increase their markups. On the other hand, higher costs might reduce the share of production workers and thereby aggregate demand. It is nevertheless possible to show that the effect on the churning intensity $\vartheta_I = \frac{\tau}{I}$ is unambiguously negative.

[13] While this result sounds similar to the findings in Aghion et al. (2001) and Aghion et al. (2005), who argue that product market competition increases growth through higher innovation incentives for incumbent firms, the mechanism is different. In my model, the ambiguous effect on the aggregate growth rate is a composition effect, whereby an increase in expansion costs (entry costs) reduces firm expansion x (entry z) but increases firms' incentives to raise markups I.

Hence, the churning intensity $\vartheta_I = \tau/I$ is still a sufficient statistic for the distribution of quality gaps. Moreover, a higher churning intensity ϑ_I decreases the distribution of markups $G(\mu;\vartheta_I) = \sum_{j=1}^{\ln\mu/\ln\lambda} \nu_j$ in a first-order stochastic dominance sense, i.e.

$$\vartheta_I^H > \vartheta_I^L \Rightarrow G\left(\mu;\vartheta_I^H\right) > G\left(\mu;\vartheta_I^L\right)$$

2. *Suppose that $p_j = \frac{1-\kappa}{\kappa}\kappa^j$, where $\kappa < 1$. Define $\theta(\kappa) \equiv \frac{1}{\ln\lambda}\ln\left(\frac{1+\vartheta_I}{1+\kappa\vartheta_I}\right)$. As in Proposition 2, the distribution of markups, the degree of misallocation $\mathcal{M}$ and the labor wedge Λ are given by*

$$G(\mu) = 1 - \mu^{-\theta(\kappa)} \qquad \text{and} \qquad \mathcal{M} = e^{-1/\theta(\kappa)}\frac{1+\theta(\kappa)}{\theta(\kappa)} \quad \text{and} \quad \Lambda = \frac{\theta(\kappa)}{1+\theta(\kappa)}.$$

The aggregate growth rate is given by $g = \ln(\lambda)\frac{I+\tau}{1-\kappa}$.

Proof. See Section A-1.6 in the Appendix. □

Proposition 5 shows that the results from the baseline model apply in a straight-forward way. It is still the case that the distribution of markups is fully determined from the churning intensity ϑ_I and that creative destruction is pro-competitive. The special case of $p_n \propto \kappa^n$ is particularly instructive. For this specification, one can show that the endogenous distribution of markups is again Pareto with shape parameter $\theta(\kappa)$. As expected, holding ϑ_I constant, the shape parameter is decreasing in κ, i.e. a more dispersed exogenous step-size distribution results in a more dispersed markup distribution in equilibrium. And as in the baseline model, an increase in ϑ_I increases the pareto tail. Because the aggregate consequences of misallocation, $\mathcal{M}$ and Λ only depend on the pareto tail of the markup distribution, the formulas from Proposition 2 directly translate. Note that the baseline model is nested as the case of $\kappa = 0$.

While these results suggest that higher values of κ induce more misallocation in equilibrium, I show below that - quantitatively - this is actually not the case. Specifically, once the model is calibrated to match the same moments as the baseline economy, the value of κ is essentially inconsequential. The intuition can be seen from the expression for the growth rate: holding I and τ constant, the growth rate is increasing in κ as the expected quality increase is larger. Hence, once the model is re-calibrated, the innovation and creative destruction intensities will adjust. I discuss this in more detail in Section 3.3 below.

2.9.2 CES-Preferences

A more fundamental assumption concerns the unitary demand elasticity. Suppose that the final good was not a Cobb Douglas aggregate but took the more general CES form

$$Y_t = \left(\int_i y_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}},$$

where y_{it} is the amount of variety i, which as before can be produced by multiple firms, i.e. $y_{it} = \sum_{f\varepsilon S_{it}} y_{fit}$. While I relegate a detailed analysis to Section OA-1.5 in the Online Appendix, I here highlight the main reasons, why the baseline case of $\sigma = 1$ simplifies the analysis and which results carry over to this more general case.

First of all, in the more general CES case, equilibrium markups are no longer necessarily determined by Bertrand competition as firms with a sufficiently large productivity advantage might prefer to charge the usual CES markup

instead of the limit price. Formally, the equilibrium markup for variety i is given by

$$\mu_i = \mu(\Delta_i) = \min\left\{\frac{\sigma}{\sigma-1}, \lambda^{\Delta_i}\right\}.$$

Secondly, while it is still the case that output can be written as $Y_t = Q_t \mathcal{M}_t L_{Pt}$, the misallocation term $\mathcal{M}_t$ and the labor wedge Λ_t now take the form

$$\mathcal{M}_t = \frac{\left(\int_i \mu(\Delta_i)^{1-\sigma}\left(\frac{q_i}{Q_t}\right)^{\sigma-1} di\right)^{\frac{\sigma}{\sigma-1}}}{\int_i \mu(\Delta_i)^{-\sigma}\left(\frac{q_i}{Q_t}\right)^{\sigma-1} di} \quad \text{and} \quad \Lambda_t = \frac{\int_i \mu(\Delta_i)^{-\sigma} di}{\int_i \mu(\Delta_i)^{1-\sigma} di}, \tag{18}$$

where the appropriate quality index Q_t is given by $Q_t = \left(\int q_i^{\sigma-1} di\right)^{\frac{1}{\sigma-1}}$. Finally, total profits in product i are given by

$$\pi(q_i, \Delta_i) \quad = \quad \left(1 - \frac{1}{\mu(\Delta_i)}\right)\mu(\Delta_i)^{1-\sigma} q_i^{\sigma-1} w^{1-\sigma} Y_t.$$

These equations highlight why the case of $\sigma = 1$ is particularly tractable. First of all, if $\sigma > 1$, the misallocation wedge $\mathcal{M}_t$ (and hence all aggregate outcomes) depend on the joint distribution of quality q and quality gaps Δ. If $\sigma = 1$, only the marginal distribution of quality gaps Δ is required. Secondly, both the level of quality q and the quality gap Δ determine the level of profits π_i and hence are state variables for the firm's dynamic programming problem. This also implies that profits (and hence also sales and employment) depend explicitly on the level of quality q so that the endogenous distribution of q is required to calibrate the model to firm-level data. In particular, the distribution of q (appropriately scaled) has to be stationary for the implied distribution of firm-level sales to be stationary. This was not required in the baseline model, as the assumption of a unitary elasticity σ implied that firm sales are directly proportional to the number of products n.[14]

It turns out that one can still make theoretical progress in analyzing this more general case. Suppose that the cost function of incumbent own-innovation is given by

$$c_t^I(I; \Delta, q) = \left(\frac{q}{Q_t}\right)^{\sigma-1} \frac{1}{\varphi_I} I^{\zeta}, \tag{19}$$

i.e. the cost of innovation scale at the same rate in q as firm profits. This requirement, which is for example also employed in Atkeson and Burstein (2010), is required to ensure that the model is consistent with Gibrat's Law, i.e. that growth rates are independent of firm size, at least for large firms.[15]

Proposition 6. *Consider the model with CES demand and let the cost of own-innovation be given by (19). Along a BGP equilibrium:*

1. *The optimal rate of own-innovation is given by a function $I(\Delta)$, i.e. is independent of q. The rate of entry z is constant and the optimal expansion rate is still given by $x = \left(\frac{\varphi_x}{\varphi_z}\frac{1}{\zeta}\right)^{\frac{1}{\zeta-1}}$.*

[14] In fact, the quality distribution in the baseline model is not stationary, even though firm-level employment, sales and profits are.

[15] The cost function in (19) does not depend on the productivity gap Δ. This is for simplicity. As long as the cost function takes the form $c_t^I(I; \Delta, q) = \left(\frac{q}{Q_t}\right)^{\sigma-1} c(I, \Delta)$, with $c(I, \Delta)$ being convex in I, one can show that the optimal rate of own-innovation is independent of q. The specification in (19) has the benefit that the resulting innovation behavior of large firms is independent of both q and Δ, i.e. consistent with Gibrat's Law.

2. *The distribution of quality gaps* $\{\nu(\Delta)\}_{\Delta=1}^{\infty}$ *is independent of* q*, stationary and given by*

$$\nu(\Delta) = \frac{\tau}{I(\Delta)} \prod_{m=1}^{\Delta} \frac{1}{1+\frac{\tau}{I(m)}}. \tag{20}$$

3. *The misallocation term* $\mathcal{M}$ *and the labor wedge* Λ *are constant and given by*

$$\mathcal{M} = \frac{\left(\sum_{\Delta=1}^{\infty} \mu(\Delta)^{1-\sigma} \nu(\Delta)\right)^{\frac{\sigma}{\sigma-1}}}{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{-\sigma} \nu(\Delta)} \text{ and } \Lambda = \frac{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{-\sigma} \nu(\Delta)}{\sum_{\Delta=1}^{\infty} \mu(\Delta_i)^{1-\sigma} \nu(\Delta)}.$$

Proof. See Section OA-1.5 in the Appendix. There I explicitly characterize the value function along the BGP, show that firms' own-innovation I is independent of q and derive (20). □

The main result of Proposition 6 is that the endogenous distributions of quality gaps Δ and productivity q are independent. The intuition is in fact simple. The optimal rate of own-innovation I is in principle a function of both state variables q and Δ. If, however, the cost function takes the form in (19), one can show that the value function is homogeneous in $q^{\sigma-1}$ and that the optimal own-innovation policy is independent of q. Moreover, the rate of creative destruction τ is constant along a BGP. This implies that the distribution of quality gaps is determined by a set of differential equations akin to (11) in the baseline model, which has the solution (20). As before, only the ratios $\{\tau/I(j)\}_j$ are required to solve for the distribution of quality gaps and hence markups. If I was constant, (20) is exactly the same solution as in the baseline model. And because markups are still fully determined from firms' quality advantage, the main economic insight from the baseline model is preserved in this more general environment: the distribution of markups is fully determined from the rate of creative destruction *relative* to firms own-innovation incentives.

The endogenous independence of productivity q and quality gaps Δ is particularly attractive because it implies that the marginal distribution of productivity q is not required to solve the model. In particular, the misallocation terms $\mathcal{M}$ and Λ follow directly from (18) and (20). These expression highlight how the demand elasticity σ affects the aggregate losses of misallocation $\mathcal{M}$. Holding $\mu(\Delta)$ and the distribution ν_Δ fixed, the aggregate costs of heterogeneous markups tend to be increasing in σ (see Hsieh and Klenow (2009)). However, two counteracting forces are also at play. First of all, a higher demand elasticity mechanically reduces markup dispersion because more products will charge a markup of $\frac{\sigma}{\sigma-1}$ holding the distribution of Δ constant. Additionally, changes in σ also affect firms' innovation and entry incentives and hence the endogenous distribution ν_Δ. In Section 3.3 I show quantitatively that the results of the baseline model are not very sensitive to the choice of σ.

3 Quantitative Analysis

I now apply this theory to plant-level data from the Indonesian manufacturing sector. This application is motivated by the recent literature on misallocation in developing countries. Because the degree of misallocation is generated endogenously, I first use the calibrated model to quantify the importance of heterogenous markups as a source of misallocation. Then I consider a counterfactual exercise and study the link between barriers to entry and monopolistic market power.

3.1 Data

The main data set for the empirical analysis is the Manufacturing Survey of Large and Medium-Sized Firms in Indonesia (Statistik Industri). This data has also been used in Amiti and Konings (2007), Blalock et al. (2008), Yang (2012) and Hsieh and Olken (2014). The Statistik Industri is an annual census of all formal manufacturing firms in Indonesia and contains information on firms' revenue, employment, capital stock, intermediate inputs and other firm characteristics.[16] I focus on the time period between 1990 and 1998, i.e. the years prior to the Indonesian financial crisis. My final sample has about 180.000 observations.[17]

The Statistik Industri data focuses on large, formal producers and therefore has a size threshold of 20 employees. In the context of a developing economy like Indonesia, this is a heavily selected sample of firms. Hsieh and Olken (2014) for example analyze data from the Indonesian economic census, which covers all producers, and find that the share of firms with less than 10 workers is essentially indistinguishable from 100 percent. At the same time, the (few) firms in the Statistik Industri data are sufficiently large to account for roughly 40% of total employment. Table 1 contains some descriptive statistics and shows that the average plant has about 140 employees. It is also the case that the firm size distribution is skewed - while the median plant has only 45 employees, the 90% quantile of the distribution is 350. Compared to the US manufacturing sector, plants in Indonesia are of course still small. In the US, one third of all establishments have more than 20 employees and such plants account for more than 90% of total employment. Moreover, the top 3.5% of plants have more than 250 employees and account for almost half of manufacturing employment.[18]

Firm size distribution				Entrants			Exiting firms		
Mean	Quantiles			Entry rate	Share of		Exit rate	Share of	
	25%	50%	90%		employment	sales		employment	sales
143	27	45	351	10.4%	5.0%	3.9%	8.2%	4.4%	3.3%

Notes: The table contains descriptive statistics on the sample of manufacturing plants in Indonesia. Columns 1 - 4 contain selected statistics about the distribution of employment. Columns 5 and 7 contain the entry and exit rate. Columns 6 and 8 report the employment share of entering and exiting firms. All results are simple averages over the time of the sample, i.e. 1991 to 1997. Table OA-1 in the Online Appendix contains the annual results.

Table 1: The Manufacturing Sector in Indonesia

For the purposes of this paper, this focus on large producers has advantages and disadvantages. On the positive side, this data covers firms for which considerations of productivity improvements and strategic pricing are more relevant. The majority of micro-firms in Indonesia are arguably subsistence entrepreneurs, which are unlikely to engage in such activities and which do not compete in the same product markets as large, formal employers.[19] Additionally, the data has a panel dimension, which allows me to use the information contained in the dynamics of markups and firm size over the life cycle to calibrate the structural parameters. To the best of my knowledge, there is no dataset covering the universe of Indonesian firms, which has a panel dimension.

The main drawback of this selection criterion is that it complicates the measurement of entry and exit as I only observe firms appearing and disappearing from the data. Table 1 shows that there are on average 10.4%

[16] To be absolutely precise, the data is collected at the plant level. As the majority of plants reports to be single branch entities, I will for the following refer to each plant as a firm.

[17] While the theory heavily exploits the fact that firms produce multiple products, my data does not contain any information at the product level. Recent empirically oriented papers that analyze product-level data include De Loecker et al. (2016), Bernard et al. (2010) or Dhyne et al. (2017). These papers show that firms routinely produce multiple products.

[18] See Table OA-2 in Section OA-2.1 in the Online Appendix for details.

[19] There is mounting evidence for the importance of "stagnant" entrepreneurs in developing economics - see e.g. Schoar (2010), Hurst and Pugsley (2012), Akcigit et al. (2015) or Hsieh and Olken (2014). That these firms do not compete in the same product markets as their larger, formal counterparts is e.g. argued in La Porta and Shleifer (2009) or La Porta and Shleifer (2014).

firms entering and 8.2% of firms exiting the data. Naturally, these firms are much smaller than the average firm so that the population of entrants (exiting firms) accounts for 5% (4.5%) of aggregate employment in the data. Interestingly, they account for an even smaller fraction of sales in the economy, reflecting the fact that they have smaller markups (as predicted by theory).

To map these moments to the theory, note that the relevant notion of firm age is the time a particular firm has been active in (potentially many) markets $i \in [0, 1]$. So if one thinks of the relevant set of product markets as the markets formal firms compete in, a new firm in the Census is indeed an entrant in the sense of the theory. I therefore consider two strategies to calibrate the model. For my benchmark calibration I treat new plants in the Census as entrants. This allows me to measure the entry-rate directly from the data. In an alternative strategy, I treat the measure of entrants as unobserved and model the empirical selection criterion by size (i.e. the size cutoff) explicitly.

3.2 The Markup Life Cycle of Indonesian Firms

The main theoretical innovation of this paper is to construct a dynamic model of firm-dynamics with endogenous markups. To calibrate the model in Section 3.3 below I therefore explicitly target the life cycle growth of Indonesian manufacturing firms. In this section I discuss in detail how I estimate this object and provide direct evidence that the variation of markups is qualitatively consistent with the theory.

Measuring Markups

To measure markups, I follow the approach pioneered by Jan De Loecker in various contributions (De Loecker et al., 2016; De Loecker and Warzynski, 2012; De Loecker, 2011) and hence relegate most of the details to the Appendix. The main benefit of this approach is that it allows me to measure firms' markups without having to take a stand on many aspects of the theory.

Because I only exploit the growth of markups long the life cycle as a calibration moment, I do not need to estimate the level of markups. This implies that I do not require an estimate of firms' production functions (or more precisely the output elasticities). To see why, consider a firm f, which is a price-taker in input markets. The optimality conditions from the firms' cost-minimization problem imply that the markup satisfies the equation

$$\mu_f = \alpha_{l,f} \times s_{l,f}^{-1}, \tag{21}$$

where $\alpha_{l,f} = \frac{\partial \ln y_f}{\partial \ln l}$ is the output elasticity of labor and $s_{l,f} = \frac{wl}{py}$ is the firm's labor share in value added (or more generally, any expenditure share of a flexible input). In my model, the output elasticity of labor is unity, so that (21) implies (2). Note that the derivation of (21) did not use any information on the structure of demand or how firms compete.[20]

If α_l was known, one could directly infer firms' markups from their observed labor shares. If α_l is not know, but assumed to be constant, (21) still identifies firms' markups up to a constant of proportionality. This is sufficient to study both the time-series and cross-sectional properties of markups. In this spirit, my baseline measure of firms' markups μ_f is the residual from the regression

$$\ln s_{l,ft}^{-1} = \delta_s + \delta_t + u_{ft},$$

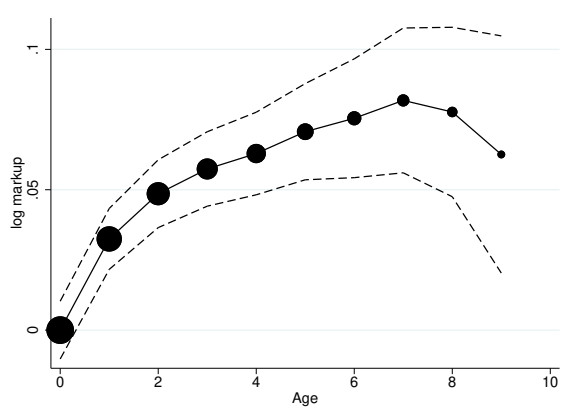
[20]Note that the allocative markup μ_{ft} depends on the payment share of *production* workers relative to sales. Empirically, I cannot precisely distinguish between production and innovation workers. The theory implies that the labor share of the entire workforce at firm f, $s_{l,f}^{Total}$, is given by $s_{l,f}^{Total} = \left(1 + \frac{w}{Y}\frac{1}{\varphi_I} I^{\zeta}\right) \mu_f^{-1} + \frac{w}{Y}\frac{1}{\varphi_x} x^{\zeta}$. Because $\frac{w}{Y}\frac{1}{\varphi_x} I^{\zeta}$ and $\frac{w}{Y} x^{\zeta}$ is constant, the variation in $s_{l,f}^{Total}$ across firms and along the life cycle is entirely driven by the variation in μ_f^{-1}.

i.e. $\ln \hat{\mu}_{ft} = \ln \hat{u}_{ft}$. Here, δ_s is a set of 5-digit industry fixed effects and δ_t is a set of year fixed effects. Under the assumption that $\alpha_{l,f}$ does not vary within 5-digit industries, the age variation of $\hat{\mu}_{ft}$ is exactly the same as if I had estimated α_l at the 5-digit level in a first stage and then calculated μ_f according to (21) using the estimated $\hat{\alpha}_l$.[21] However, I also consider richer specifications, where I explicitly control for firms' input choices like the capital- or material intensity, to allow for additional variation in output elasticities across firms within 5-digit industries. Furthermore, I also report results where I measure markups from firms' material shares instead of labor shares.

Results

A key prediction of the theory is that markups should increase in age, at least for the majority of firms. In Figure 3, I show to what extent this is the case in the Indonesian data. I want stress that this pattern is estimated from the time-series variation and not from the cross-sectional age-size relationship. More specifically, I focus on all firms that entered the data after 1990 (which allows me to measure plant-age in a consistent way) and then calculate the average markup by age relative to entering firms. As in the theory, there is attrition as firms exit the market and the size of the dots reflects the size of the surviving firms in the cohort. This schedule therefore refers exactly to the expression characterized in Proposition 3 and displayed in Figure 1 and I will use it as an explicit moment to calibrate the model.

As predicted by theory, markups increase in age for young firms. In particular, they show a somewhat concave profile and seem to level off around age 8 (even though markups for old firms are not very precisely estimated). Quantitatively, markups of 7 year old firms are on average 8% larger compared to recent entrants. Through the lens of the theory this implies that firms engage in own-innovation activities. If firms were to only grow horizontally by adding new markets as in Klette and Kortum (2004), markups and age should not be systematically related.



Notes: The figure shows the life cycle of markups. To calculate the markup life cycle, I focus on the unbalanced panel of firms entering the economy after 1991. I calculate log markups within 5-digit-industry-year cells, then calculate the average by the age of the cohort and normalize log markups of entering cohorts to zero. Because of attrition, the size of the cohort is declining in age. The dots reflect the size of the cohort. I also depict the 90% confidence intervals around the estimated average profile.

Figure 3: The Life Cycle of Markups in Indonesia

To study the life cycle profile of markups more systematically, Table 2 contains additional regression evidence for the patterns in Figure 3. I focus on regressions of the form

$$ln\left(\mu_{ft}\right) = \delta_t + \delta_s + \beta \times age_{ft} + \varrho \times ln\left(k_{ft}/l_{ft}\right) + h'_{ft}\gamma + u_{ft}, \tag{22}$$

[21] Furthermore, my data is standard in the sense that it does not contain information on firm-specific prices. Hence, to estimate the output elasticity θ (which corresponds to *physical* output) one needs to impose additional structure.

where k/l denotes the firms' capital-labor ratio, h contains additional firm-characteristics and δ_t and δ_s denote year and 5-digit industry fixed effects. As in Figure 3 I focus on the unbalanced panel of firms entering the economy after 1990. The results are contained in Table 2.

	Dependent variable: log markup						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	0.0137***	0.0103***	0.0106***		0.0134***	0.0108***	0.0237***
	(0.00122)	(0.00117)	(0.00116)		(0.00158)	(0.00127)	(0.00185)
Entry				-0.0250***	0.00777		
				(0.00638)	(0.00744)		
Exit				-0.0176**	-0.0144*		
				(0.00835)	(0.00836)		
Industry FE	✓	✓	✓	✓	✓	✓	
Year FE	✓	✓	✓	✓	✓	✓	
Firm FE							✓
Control for capital-intensity		$\ln k/l$	non-param.	$\ln k/l$	$\ln k/l$	$\ln k/l$	
N	55212	55212	55212	48556	48556	42434	14779
R^2	0.219	0.287	0.305	0.293	0.294	0.298	0.672

Notes: Robust standard errors in parentheses with ***, **, and * respectively denoting significance at the 1%, 5% and 10% levels. I focus on the unbalanced panel of firms, who enter the market after 1990. I use the data from 1991 to 2000. $ln\,(k/l)$ denotes the (log) capital-labor ratio at the firm level. 'Entry" and "Exit" are indicator variables for whether the firm enters (exit) the market in a given year. In column 6 I focus on the balanced panel, i.e. only consider firms that survive to the end of my sample period. Industry fixed effects control for industry affiliation at the 5 digit level.

Table 2: The Life Cycle of Markups in Indonesia

The first column contains the specification displayed in Figure 3. On average, markups increase by roughly 1.4% per year. In columns 2 and 3, I include firms' capital-labor ratio to control for a correlation between capital-intensity and firm size (and hence age). Doing so reduces the estimated age-coefficient slightly. In column 2 I simply include the log of firms' capital-labor ratio, in column 3 I control in a non-parametric way by including 50 fixed effects for 50 quantiles of the distribution of capital-labor ratios. In column 4, I show that both entrants and exiting firms have lower markups. This is consistent with the model: both entering and exiting firms are small and therefore - on average - young.[22] In column 5 I include firm age directly, which should account for the low markup of both entrants and exiting firms. For the case of entrants this is indeed borne out by the data. The results for exiting firms are not consistent with theory. In the model exit is *only* a function of the number of markets and not the age of the firm. However, the model implies that markups are decreasing in size holding age fixed. Hence, conditional on age, exiting firms should actually have higher markups. While the negative coefficient does increase once age is controlled for, exiting firms still have lower markups given their age. In column 6 I directly control for selection by conditioning on survival. In the theory, there is no selection in that the distribution of markups conditional on age is the same for all firms. In the data, the growth of markups shown in Figure 3 could stem from a higher exit hazard of firms, that systematically have low markups. Column 6 shows that markups are increasing over time even for those firms that do survive until the end of the sample. Finally, in column 7, I focus on the cohort of firms who entered the economy in 1991 and estimate the age profile of markups by including a full set of firm fixed effect. Again I find a positive coefficient, which if anything is slightly larger than my baseline estimate.[23]

[22]See also Foster et al. (2008), who - using US data - report qualitatively similar findings for particular industries selling homogenous products.

[23]In Section OA-2.6 in the Appendix I present additional robustness checks for these results. In particular, I show that the results do not substantially depend on whether or not I correct the measure of markups for measurement error as suggested in De Loecker and Warzynski (2012). I also consider the case of material shares in total sales. Markups should distort all factors within the firm

Alternative misallocation frictions My theory abstracts from any distortions to firms' input choices. This allows me to use the data on firms' factor shares to measure markups. This interpretation might be misleading if firms are subject to frictions, which distort their input choices. Because I am only use the growth of markups over the firms' life cycle, input distortions, which are constant at the firm-level do not invalidate the results in Table 2. More problematic are frictions, which systematically change in firm age or size (see e.g. Guner et al. (2008), Buera and Fattal Jaef (2016) or Bento and Restuccia (2017)). While firm-specific market power is a natural mechanism to generate such "size-dependent policies", there could be other possibilities for why large firms appear to be constrained - see e.g. the discussion in Hsieh and Olken (2014) and David and Venkateswaran (2019), who provide a methodology to identify different sources of misallocation.

In Section OA-2.6 in the Appendix, I present additional evidence on the comparison between factor shares reflecting markups or input distortions. In particular, I show that exporters and firms relying on FDI have in fact higher revenue productivity and that firms, who report to be capital-constrained, have lower revenue productivity. Hence, as often found in the literature, typical performance measures are positively correlated with marginal products. While this correlation is natural in environments where factor shares reflect market power, it is somewhat harder to rationalize with standard models of for example credit-constraints.[24]

3.3 Markups and Misallocation in Indonesia

To quantify the efficiency losses from market power I now calibrate the model to moments from the Indonesian firm level data and calculate the aggregate costs of misallocation. Importantly, I do *not* assume that the entirety of the variation in labor shares is due to market power. In contrast, I use the estimated life cycle growth of markup as a moment to calibrate the model and then calculate the losses from misallocation from the equilibrium distribution of markups as implied by the theory.

Calibration

The model is very parsimonious. Given a rate of time preference ρ, which I set exogenously, the theory is fully parametrized by five parameters: the innovation step-size λ, the cost shifters for innovation, incumbent creative destruction and entry φ_I, φ_z and φ_x and the curvature of the innovation and expansion technology ζ.

My calibration strategy is as follows. As shown in the expression for the life cycle of firm size and markups, all micro-moments related to the process of firm-dynamics only depend on the three endogenous outcomes (I, x, τ) and the exogenous step size λ. Hence, the model provides a direct mapping from the data to (I, x, τ) and λ and this mapping does not depend on ζ nor ρ. I then use the equilibrium conditions to find the required structural parameters to yield (I, x, τ) as equilibrium outcomes consistent with optimal behavior and market clearing. For a given cost elasticity ζ, the uniqueness of the equilibrium implies that there is a unique mapping from the policy functions (I, x, τ) to the structural parameters $(\varphi_I, \varphi_z, \varphi_x)$. Credibly identifying the curvature parameter ζ is difficult without exogenous variation in innovation costs. I therefore follow Acemoglu et al. (2016) and assume that $\zeta = 2$ for my baseline results and provide robustness.

equally, so that (21) should also hold for $s_{M,ft} = \frac{m_f}{py_f}$, i.e. the share of sales going to materials. The results are very similar to the ones reported in Table 2 when firm's material-labor ratio $\ln(m/l)$ is controlled for. This is important, as there is a strong positive correlation between $\ln(m/l)$ and age, i.e. older firms rely relatively more on materials. If this change in production factors is not controlled for, the correlation between material shares and age is negative.

[24] Benchmark dynamic models of financial constraints imply exactly the opposite pattern: borrowing constraints tend to bind in the early stages of the life cycle and get relaxed as the firms ages (see e.g. Clementi and Hopenhayn (2006)). However, in models of financial credit constraints, where firms face productivity shocks (see e.g. Buera et al. (2011), Moll (2014) or Midrigan and Xu (2014)), the relationship between factor shares and firm age is less clear as old firms had more time to accumulate savings but might also have experienced a sequence of favorable productivity shocks, which increase their borrowing needs.

Targets	Data	Model	Calibrated parameters			Endogenous outcomes	
life cycle of Markup	0.082	0.082	φ_I	Productivity of innovation	8.896	I	0.561
life cycle of Employment	0.494	0.494	φ_x	Productivity of expansion	1.077	x	0.207
Entry rate	10.4%	10.4%	φ_z	Productivity of entry	2.590	z	0.032
Aggregate productivity growth	3%	3%	λ	Innovation step size	1.038	τ	0.240
-	-	-	ζ	Curvature of cost function	2		
-	-	-	ρ	Rate of time preference	0.05		

Notes: This table contains the calibrated parameters, the targeted data moments, the resulting moments in the model and the equilibrium outcomes for I, x and z. Given (I, x, z) all other equilibrium allocations can be calculated. I measure the life cycle of markups and employment as the log difference between entrants and seven year old firms - see the left panel of Figure 3 for the case of markups and Section OA-2.6 in the Online Appendix for the case of employment. The entry rate stems from Table 1. The two parameters ζ and ρ are set exogenously. As explained in the text, the mapping from the data moments to (I, x, z, λ) does not depend on ζ and ρ. See Section A-2.2 in the Appendix for details.

Table 3: Calibration

To identify the four structural parameters $(\varphi_I, \varphi_z, \varphi_x, \lambda)$ I use four moments. All of these moments have closed form expressions in the theory. First of all, I target the life cycle of markups, i.e. the average markup of 7 year old firm relative to entrants, given in (15). Secondly, I match the observed life cycle of employment, i.e. average employment of 7 year old firms relative to entrants, given in (17). Third, I target the entry rate, which is given by (see Section 2.7)

$$\text{Entry Rate} = \frac{z}{F} = \frac{z}{\frac{1-\vartheta_x}{\vartheta_x} \times \ln\left(\frac{1}{1-\vartheta_x}\right)} = \frac{x}{\ln\left(\frac{z+x}{z}\right)},$$

where recall F denotes the number of firms in equilibrium. Finally, I match a given rate of aggregate productivity growth $g = \ln \lambda (\tau + I)$. These theoretical relationships allow me to write the four targeted moments directly in terms of the four unknowns (I, τ, z, λ). Calibrating the model therefore reduces to solving four non-linear equations.

Table 3 reports the results of this exercise. In terms of data moments, I require the model to match the fact that markups increase by 0.08 log points over a 7 year horizon (displayed in Figure 3) and that firms in Indonesia increase their employment by roughly 0.5 log points (i.e. a factor of 1.6) in the first 7 years of their life (see Figure 4 below, where I depict the employment life cycle from both the data and the model). To identify the flow rate of entry z, I require the model to be consistent with the observed entry rate of 10.4% reported in Table 1. Finally, I discipline λ to match an aggregate rate of productivity growth of 3%.[25] In Table 3 I report the data and model moments, the structural parameters and the implied endogenous innovation outcomes.

To further illustrate the mapping between the structural parameters, the equilibrium outcomes and the implied moments, Table 4 contains a sensitivity matrix and reports the change in equilibrium outcomes and moments for a 5% increase in the respective structural parameters. The results conform well with the economic intuition of the theory. First of all, an increase in the efficiency of innovation (φ_I), expansion (φ_x) and entry (φ_z) raises own-innovation I, incumbent creative destruction x and entry z respectively. Furthermore, there is a sizable extent of

[25] A detailed description of the construction of all these data moment is contained in Section A-2.2 of the Appendix. There I also present additional regression evidence for the life cycle of employment. In particular, note that in contrast to firms in India, which - according to Hsieh and Klenow (2014) - experience essentially no growth as they age, Indonesian firms do grow over time conditional on survival. However, both the sample of firms and the methodology underlying my calibration is different. First of all, the Indonesian data is biased towards bigger, formal firms. Secondly, my moments are estimated from panel data and not inferred from the cross-sectional age-size relationship. This turns out to be important as the cross-sectional age-size relationship in Indonesia is also relatively flat, which could be due to measurement error in firm age. See Section OA-2.6 in the Appendix, where I replicate the cross-sectional age-size relationship using the methodology by Hsieh and Klenow (2014). This difference between the "true" life cycle and the cross-sectional patterns do not seem to be unique to the Indonesian context, but are also present in for example Chile (see Buera and Fattal Jaef (2016))

	Change in ...				Initial
	φ_I	φ_x	φ_z	λ	level
Effects on endogenous outcomes					
Rate of own innovation (I)	4.0%	-1.1%	-0.9%	1.9%	0.561
Incubment creative destruction (x)	0.0%	5.0%	-4.8%	0.0%	0.2078
Entry (z)	8.1%	-21.9%	38.1%	24.3%	0.0326
Creative destruction (τ)	1.1%	1.4%	1.1%	3.3%	0.2404
Effects on equilibrium moments					
Life cycle of markups	3.6%	-3.3%	0.6%	5.8%	0.0818
Life cycle of employment	-1.5%	6.9%	-7.9%	-3.5%	0.494
Entry rate	3.5%	-7.1%	12.9%	10.2%	0.104
Growth rate	3.1%	-0.3%	-0.3%	7.3%	0.03

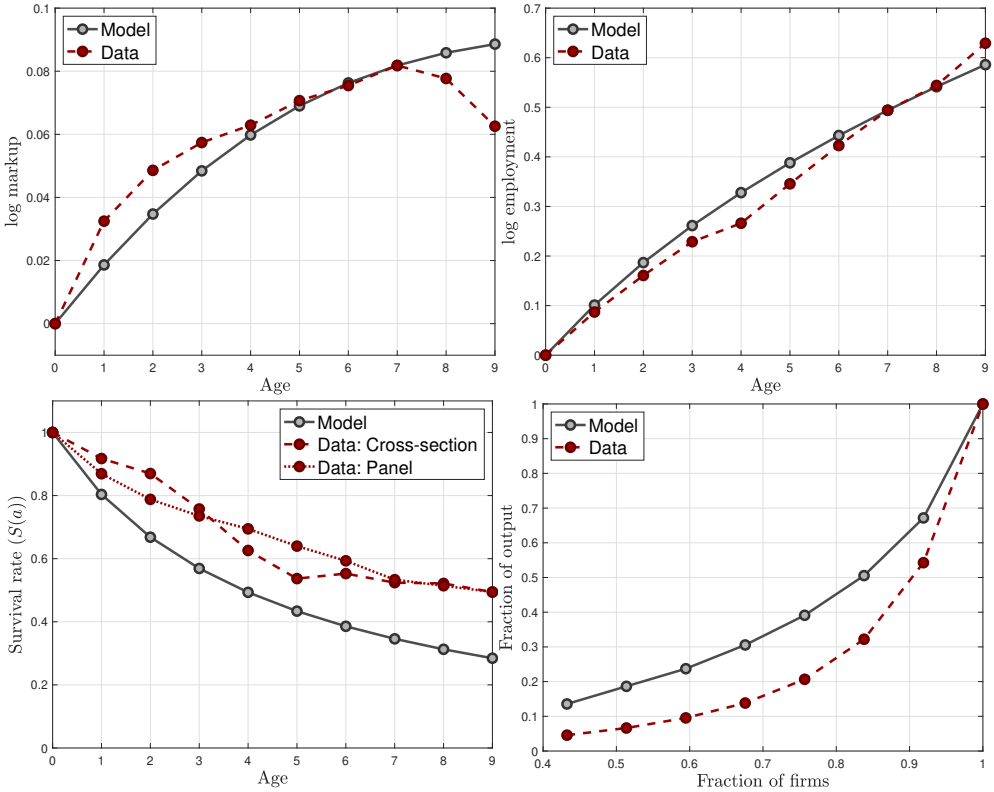
Notes: The table reports the effect of a 5% change in the relative innovation efficiencies $(\varphi_I, \varphi_x, \varphi_z)$ and the quality increase $(\lambda - 1)$ on the endogenous outcomes (top panel) and the equilibrium moments (lower panel).

Table 4: Sensitivity Matrix

crowd-out in that an increase in φ_x reduces the equilibrium amount of entry z and an increase in φ_z lowers incumbent creative destruction x. Similarly, both of these parameters also negatively affect incumbent own innovation, as the higher rate of creative destruction increases the effective discount rate of existing firms. That neither φ_I nor λ affect the equilibrium level of creative destruction by incumbents is apparent from (9). Note that a higher efficiency of own-innovation φ_I increases the amount of entry by increasing the value of firms.

The implied changes in the resulting moments are reported in the lower panel of Table 4. These are consistent with the theoretical life cycle patterns shown in Figure 2. An increase in the efficiency of own-innovation φ_I increases markup growth and reduces employment growth. The latter is the combination of both rising markups and lower sales growth as creative destruction τ rises and the expansion rate x is unaffected. Increases in the efficiency of incumbent creative destruction φ_x reduce the extent of life cycle markup growth as firms add new, low-markup products to their portfolio at a faster rate. In contrast, higher entry efficiency *increases* markup growth, despite lowering incumbent own innovation I and increasing creative destruction τ. The reason is that it also discourages creative destruction of incumbents. The effect of the step size λ is also intuitive: the rate of growth, the entry rate and the extent of markup growth increase and the slope of the employment life cycle declines.

Non-targeted moments To get a better sense of the link between the model and data, Figure 4 contains a set of outcomes, which I did not explicitly target. In the top two panels I depict the calibrated and observed life cycle patterns for markups (left panel) and employment (right panel). While the model is only calibrated to match the data for 7 year old firms, it captures the general age pattern for markups and employment well. In the lower panels of Figure 4 I focus on two additional predictions, namely the dynamic patterns of exit (left panel) and the cross-sectional size distribution (right panel). In equation (16) I derived the theoretical prediction for the probability of survival, $S(a)$. Through the lens of the model, which is stationary, I can measure $S(a_f)$ either from the cross-section at time t and "count" the firms of different ages, or by following a cohort through its life cycle. In the lower left panel of Figure 4, I depict the model's implication for $S(a)$ and the corresponding data - both measured from the cross-section in 2000 and from the evolution of the cohort, which entered in 1991. Two properties stand out. First of all, the survival probabilities estimated from the cross-section and from the panel turn out to be quite similar. Secondly, the model captures the dynamics of exit relatively well, even though it slightly over-predicts the extent of "shake-out", i.e. firms exit at too fast a rate. The bottom right panel displays the fraction of aggregate output, which is accounted for by the smallest x% of firms, i.e. the Lorenz curve of the firm size distribution. Both



Notes: The figure displays the model's prediction and the data for the life cycle of markups (top left panel), employment growth (top right panel), survival (bottom left panel) and the concentration of value added (bottom right panel). Firm survival is measured both by the share of firms by age in the 2000 cross-section (relative to the number of firms at the time of entry) and the share of firms of the entering cohort in 1991 by age. For the employment life cycle see Section A-2.2 in the Appendix. To measure the distribution of firm size, I drop the highest and lowest 3 percent of firms to account for the possibility of measurement error.

Figure 4: Model vs Data: Markup and Employment Dynamics, Survival and Concentration

Markups and Misallocation					
θ	$E[\mu]$	$\sigma(\ln \mu)$	$\sigma(\ln \mu_f)$	$\mathcal{M}$	Λ
9.5	11.8%	0.103	0.079	0.995	0.9

Notes: The table reports the endogenous tail parameter of the markup distribution θ, the average markup ($E[\mu]$), the dispersion of log markups across products ($\sigma(\ln \mu)$) and firms ($\sigma(\ln \mu_f)$) and the two misallocation wedges $\mathcal{M}$ and Λ (see (3) and (4)).

Table 5: Markups and Misallocation in Indonesia

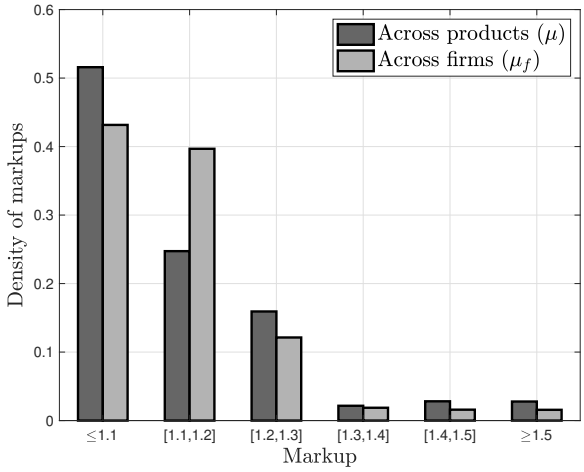
in the model and the data the curve is below the 45-degree line, reflecting the dispersion in firm size. The empirical distribution is, however, more unequal than its theoretical counterpart, i.e. the model under-predicts the extent of concentration.[26]

The theory also makes predictions about the correlation of markups and size. While the model predicts that markups are increasing in firm size, the relationship is quantitatively small compared to the data. In Section A-2.3 in the Appendix, I derive an analytic expression for the average markup as a function of size, i.e. the size-based analogue to Proposition 3. For the calibrated parameters, the implied elasticity between markups and size is positive, but very small: 0.006. Empirically, the elasticity is also positive but much larger: 0.23. The reason why the model under-predicts the extent to which markups rise with size is related to the discussion around Figure 2: the only reason why larger firms should have higher markups, is that they are on average older. However, holding age fixed, the correlation between markups and size is negative as successful expansion draws new, low-markup products into the firm. In the calibrated model, this composition effect is strong enough to almost undo the positive correlation induced by the co-movement in the life cycle of markups and employment. This suggests an important role for systematic heterogeneity across producers as the correlation between size and markups would, for example, be stronger if some firms were both more efficient to expand and would also have systematically higher quality draws, which would allow them to charge higher markups.

Implications for Markups and Misallocation

I now turn to the aggregate implications of the calibrated model, which I report in Table 5. The equilibrium churning intensity $\vartheta_I = \frac{\tau}{I}$ implied by Table 3 is equal to 0.43. Proposition 2 therefore implies that the distribution of markups across products is pareto with shape $\theta = 9.5$. Average markups are therefore 12%. This in the range of the estimates of De Loecker and Warzynski (2012), albeit at the lower end. Depending on the specification used, they estimate markups between 10% and 28%. They are also markedly lower than the results reported in De Loecker and Eeckhout (2017), who report a sales-weighted average markup of 20% in 1980 and 60% in 2012 for publicly traded firms in the US. In my model, average-markups are directly implied from the process of firm-dynamics and a given aggregate growth rate. Intuitively: for average markups to be higher (given the estimated life cycle slope) the step size λ would need to be higher. This however, would also imply a higher aggregate growth rate. In terms of markup dispersion, the calibrated model implies a standard deviation of log markups of about 0.1. While this dispersion is the welfare-relevant measure of markup heterogeneity, it does not directly compare to empirical measures of TFPR dispersion across firms. If firms are active in multiple product markets, the dispersion across firms is lower than the overall heterogeneity faced by consumers. This discrepancy is seen in Figure 5, where I depict the equilibrium distribution of markups across products (dark bars) and firms (light

[26] In the model, firms' *only* margin of employment growth is to enter in new markets and to replace other producers - productivity growth in existing markets actually reduce employment through increasing markups. This sharp distinction is conceptually and analytically useful. It is, however, restrictive. For example, if the elasticity of demand exceeded unity, increases in quality would also led to increases in employment and the model could rationalize a given slope of the age-employment schedule with margins other than creative destruction. See also Garcia-Macia et al. (2016) and Luttmer (2010).



Notes: The figure shows the distribution of markup at the product level (dark bins) and at the firm level (light bins). The results are based on the calibration reported in Table 3.

Figure 5: The Stationary Distribution of Markups\

bars). It is clearly seen that the markup distribution across firms is compressed because it neglects the dispersion of markups within firms. Quantitatively, the dispersion across firms underestimates the actual dispersion by about 20%. The static macroeconomic consequences of firms' market power are summarized by $\mathcal{M}$ and Λ. The model implies that TFP is lowered by 0.5% and wages are depressed by 10% relative to their social marginal product. The implied reduction in TFP seems small, especially compared to the much bigger numbers reported in Hsieh and Klenow (2009) or De Loecker and Eeckhout (2017). There are two reasons. First of all, the model implies that that markups only account for a small fraction of the observed dispersion in revenue products. Empirically, the standard deviation of log labor shares is 0.74, which is consistent with Hsieh and Klenow (2009) who find numbers of around 0.7 in China and India. The remainder could hence be explained by other frictions, adjustment costs, model misspecification or measurement error.[27] Secondly, Hsieh and Klenow (2009) consider an elasticity of substitution of three, whereas I impose a unitary demand elasticity. Recall that the change in aggregate TFP is approximately given by $d\ln TFP = -\frac{\sigma}{2} d\text{var}\left(\ln TFPR\right)$, where σ is the elasticity of substitution across varieties. For $\sigma = 1$ and $d\text{var}\left(\ln TFPR\right) = 0.103^2$, one exactly recovers a TFP loss of 0.5% as reported in Table 5.[28]

The model also has implications for the sources of productivity growth and reallocation. The model implies that entrants account for $\frac{z}{\tau} \approx 13.5\%$ of aggregate creative destruction. In terms of productivity growth, the calibrated model implies that the share of aggregate growth accounted for by firms' own-innovation is given by about 70%. This is similar to Garcia-Macia et al. (2016), who estimate that about 75%-80% of incumbent growth is due to own-quality improvements. Note however, that own-quality improvements in Garcia-Macia et al. (2016) increase employment (as markups are assumed to be constant), while such productivity increases in my model are fully reflected in firms' markups. These two facts imply that the share of aggregate growth accounted for by entering firms, $\frac{z}{I+\tau}$, is low - it is only 4%. There are two main reasons why this is the case. First of all, I abstracted from the entry of new varieties. If new varieties are more likely to be produced by entering firms, the entry share in aggregate growth could increase. Secondly, I restrict the step size of innovation, λ, to be the same across all sources.

[27] David and Venkateswaran (2019) also find that the dispersion in markups accounts for a small share of the dispersion in the marginal product of capital across Chinese firms.

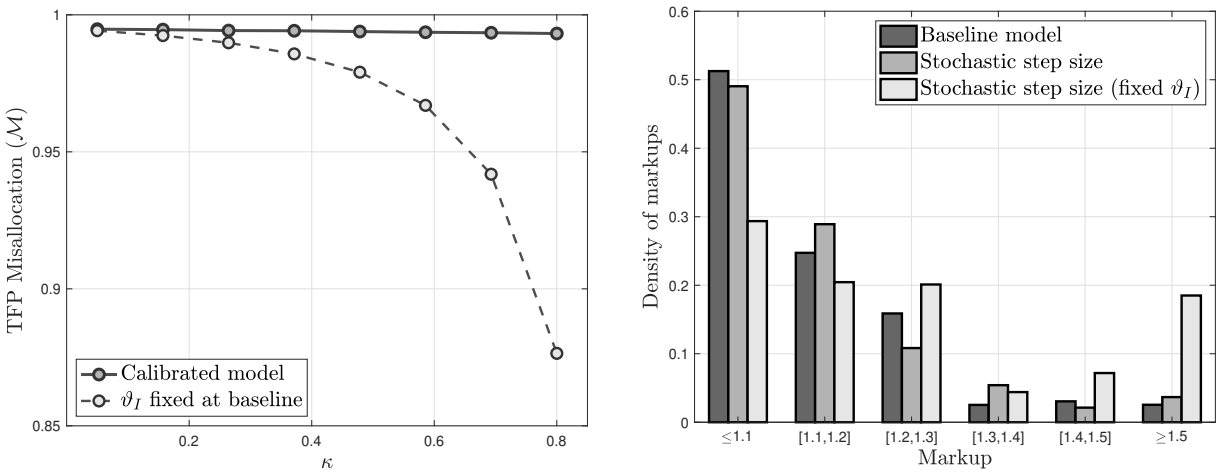
[28] The formula $d\ln TFP = -\frac{\sigma}{2} d\text{var}\left(\ln TFPR\right)$, used in Hsieh and Klenow (2009), relies on the assumption of physical productivity and $TFPR$ be to independent and log-normally distributed. They estimate that $\text{var}\left(\ln TPFR^{IND}\right) - \text{var}\left(\ln TPFR^{US}\right) \approx 0.24$ For $\sigma = 3$, this implies a loss in aggregate TFP of 36%.

If entrants were to enter with technologies, which represented a drastic innovation, a given entry rate could be consistent with a larger share of growth.

Stochastic Step Size and CES Preferences: Quantitative Implications

In Section 2.9 I showed how to generalize the theory along two dimensions: I allowed for the step size of successful innovations to be stochastic and I characterized the economy with CES preferences. In this section I show that these extension do not significantly change the quantitative magnitude of the losses from misallocation. For brevity I focus on the consequences for $\mathcal{M}$. The results for the labor wedge Λ are similar.

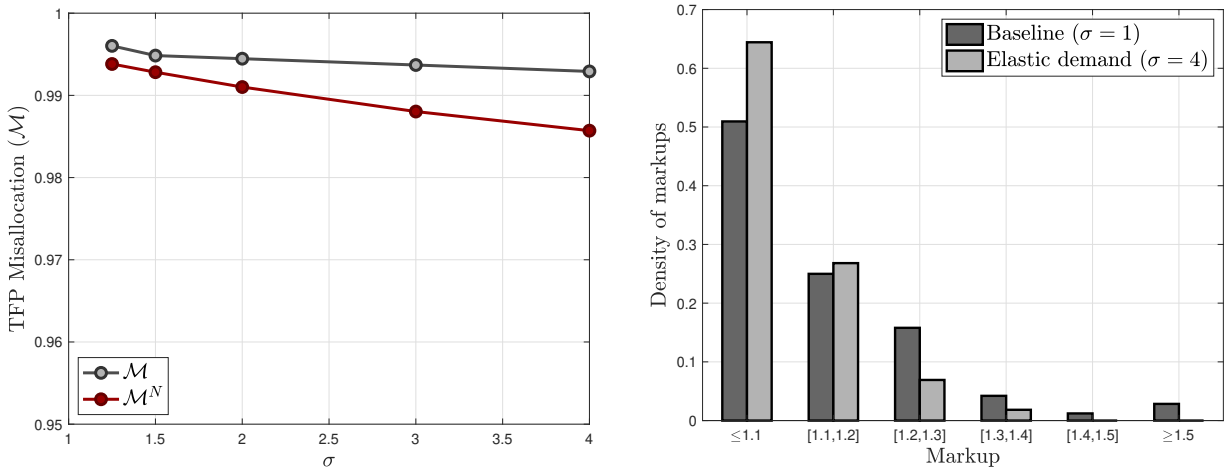
Stochastic Step Size To see that the restriction of a common step size is not restrictive, consider the model where the step size is drawn from $p_n = \frac{1-\kappa}{\kappa}\kappa^n$ (see Proposition 5). The left panel of Figure 6 shows the extent of misallocation $\mathcal{M}$ as a function of κ. The solid line correspond to the case where - for every κ - I recalibrate all parameters to match the exact same moments as in my baseline calibration. It is clearly seen that the line is essentially flat. Hence, as far as the aggregate implications for the degree of misallocation are concerned, heterogeneity in the number of quality steps is not particularly important. This is also seen in the right panel of Figure 6, which contains the endogenous distribution of markups and shows that the distribution for the stochastic step size model is very close to the baseline calibration.



Notes: The left panel displays the misallocation wedge $\mathcal{M}$ as a function of κ. The implications for the re-calibrated model are shown as the solid line. For comparison I also depict the case where I vary only κ, but keep ϑ_I at its baseline value. In the right panel I show the stationary distribution of markups in the baseline model (dark bars), in the model with a stochastic step size for case of $\kappa = 0.5$ (medium bars) and when I keep ϑ_I at its baseline value (light bars).

Figure 6: Misallocation and Markups: Stochastic Step Size

Recalibrating the model is crucial to reach this conclusion as the parameter κ obviously has a mechanical effect on the distribution of markups holding the churning intensity ϑ_I fixed. When I fix ϑ_I at its baseline value, misallocation increases substantially and aggregate TFP is reduced by more than 10% for $\kappa = 0.8$. The reason is, of course, that the distribution of markups becomes much more dispersed when firms' quality steps are stochastic but ϑ_I is takes an parametric. In particular, as seen in the right panel, the share of products with markups exceeding 50% increases markedly.



Notes: The left panel shows the misallocation wedges $\mathcal{M}$ and $\mathcal{M}^N$ for different values of the demand elasticity σ. In the dark line I show the actual expressions when the model is calibrated to the same moments as the baseline model. In the red line I depict the "naive" measure, which keep the distribution of markups at their baseline level. The right panel shows the distribution of markups across products for both the baseline economy and for the economy with $\sigma = 4$.

Figure 7: Misallocation in the Model with CES Preferences

The CES Model The restriction to a unitary elasticity of substitution is potentially more consequential. It turns out, however, that the implications of the baseline model for the costs of misallocation are also quantitatively robust to changes in the demand elasticity.[29] Proposition 6 established that the misallocation wedge along the BGP is given by

$$\mathcal{M}(\sigma) = \frac{\left(\sum_{\Delta=1}^{\infty} \mu(\Delta)^{1-\sigma} \nu_{\Delta}\right)^{\frac{\sigma}{\sigma-1}}}{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{-\sigma} \nu_{\Delta}} \quad \text{where } \mu(\Delta) = \min\left\{\frac{\sigma}{\sigma-1}, \lambda^{\Delta}\right\},$$

where the notation makes the dependence on σ explicit. This expression highlights that the demand elasticity σ has three effects. First of all, holding the distribution of quality gaps ν_{Δ} and the corresponding markups $\mu(\Delta)$ fixed, the demand elasticity σ determines how this heterogeneity is correctly aggregated. Secondly, σ directly determines the mapping from quality gaps Δ to the markups firms actually post. The higher σ, the lower the optimal CES-type markup $\frac{\sigma}{\sigma-1}$, i.e. a higher demand elasticity truncates the right tail of the markup distribution. Finally, the distribution of quality gaps ν_{Δ} itself is endogenous and changes as a function of σ.

Focus first on the mechanical effect of the correct aggregator. In particular, taking the distribution of markups from the baseline model as given (i.e. $G(\mu) = 1 - \mu^{-\theta}$, where $\theta = \frac{\ln(1+\vartheta_I)}{\ln \lambda}$ stems from the baseline calibration), the misallocation wedge is given by

$$\mathcal{M}^N(\sigma) = \frac{\left(\int \mu^{1-\sigma} dG(\mu)\right)^{\frac{\sigma}{\sigma-1}}}{\int \mu^{-\sigma} dG(\mu)} = \frac{\theta+\sigma}{\theta}\left(\frac{\theta}{\theta+\sigma-1}\right)^{\frac{\sigma}{\sigma-1}}.$$

I refer to this measures as the "naive" measure, because it abstracts from any feedback of the demand elasticity to the markups firms actually post. This expression is nevertheless helpful because it turns out that $\mathcal{M}^N$ is a relatively tight upper bound bound for the full effects captured by $\mathcal{M}(\sigma)$. In the left panel of Figure 7 I depict this naive measure in red. These naive losses from misallocation are increasing in σ. Quantitatively, this mechanical effect can potentially increase the TFP cost of misallocation by about 1 percentage point.

The actual misallocation losses $\mathcal{M}(\sigma)$, which are depicted in the dark lines in Figure 7, are smaller. For a

[29] Section OA-1.5 in the Appendix contains all detailed derivations for the CES model.

demand elasticity of 4, markups can reduce TFP by about a percentage point, i.e. the model "adds" about half a percentage point of TFP losses relative to the baseline model. That the misallocation losses implied by the naive measure are an upper bound for the actual losses is intuitive. By truncating equilibrium markups at $\frac{\sigma}{\sigma-1}$, a higher demand elasticity tends to lower misallocation by reducing the dispersion of markups. Additionally, it turns out that the endogenous distribution of quality gaps ν_Δ is such that there is more mass on low markup products. This is for example seen in the right panel of Figure 7, where I depict the stationary distribution of markups for the baseline model and for the case of $\sigma = 4$, where the highest markup in the economy is 33%. Note that there is no mass on the set products in the last two bins of the histogram and that the whole markup distribution is shifted to the left. Hence, both the level of markups and their dispersion is lower. That the costs of misallocation are nevertheless slightly higher than in the baseline model is due to the CES aggregator.

3.4 Creative Destruction: The Role of Expansion and Entry Costs

How important are frictions like expansion costs for existing firms or entry costs for new producers for the process of creative destruction? And what are the implications for the aggregate losses of misallocation, the economy-wide rate of productivity growth and the endogenous distribution of firm size? In Proposition 4 I characterized these effects qualitatively. I now use the calibrated model and data from the US to provide a quantitative answer. I focus on a comparison of the stationary equilibria (or balanced growth paths) and abstract from transitional dynamics.

To discipline the counterfactual change in entry and expansion costs, I target two moments which have a natural mapping to these unobservables: the rate of entry and the extent of life cycle employment growth. More specifically, I start from the calibrated economy above and recalibrate the expansion costs $\left(\varphi_x^{-1}\right)$ and the entry costs $\left(\varphi_z^{-1}\right)$ to match the entry rate and the life cycle growth rate of employment of manufacturing firms in the US. All the remaining parameters are left unchanged. If these were indeed the only differences between the US and Indonesia, the resulting estimates would be informative about the differences in misallocation and growth between these two economies. As an expositional shorthand I will therefore refer to the recalibrated economy as "the US".

As a benchmark for the rate of life cycle growth in the US, I build on the results of Hsieh and Klenow (2014), who report that US firms grow by a factor 2 at the 10 year horizon. This corresponds to an annual rate of employment growth of about 7% conditional on survival. As for the entry rate, I target a value of 8% for the baseline results, which is consistent with Karahan et al. (2015), who report a start-up rate of between 8% and 11% for the whole economy and Akcigit et al. (2015), who calculate an entry rate of 7.5% in the US manufacturing sector. Hence, the US entry rate is lower and the rate of life cycle growth higher.

The results of this exercise are contained in Table 6. In the top panel, I report the two new moments and the resulting estimates for the entry and expansion costs. I express these estimates relative to the values for the Indonesian economy reported in Table 3. While both entry and expansion costs are lower in the US, the expansion margin is particularly important. The entry technology in the US is about 15% more productive, the costs for existing firms to break into new markets are about a third lower. The intuition for these results is simple. For the model to generate a faster rate of life cycle employment growth, firms need to be willing to expand more aggressively. This calls for lower expansion costs in the US. This, however, reduces the equilibrium entry rate. In a stationary equilibrium, the mass of entrants has to be equal to the mass of exiting firms, i.e. the mass of single-product firms experiencing a creative destruction shock. Even though expansion costs increase the rate of creative destruction, there is still less exit, simply because in an economy populated by large firms it is less likely that a producer who gets replaced in a particular product exits. To match the observed entry rate of 8%, more entry is required. Entry barriers in the US therefore also have to be lower. In the remaining panels I report the equilibrium implications. I focus on three sets of results, which concern the distribution of firm size (Panel A), the distribution of markups

		Heterogenous Markups			Constant Markups
		Indonesia	US	Change	Change
Calibration					
Entry Rate		10.4%	8%		
Employment life cycle		1.7	2		
Expansion Costs	$\left(\varphi_x/\varphi_x^{IND}\right)^{-1}$	1	0.67	-33%	-21%
Entry Costs	$\left(\varphi_z/\varphi_z^{IND}\right)^{-1}$	1	0.86	-14%	-2%
Equilibrium Implications					
Panel A: The Distribution of Firm Size					
Number of firms	F	0.313	0.122	-60.1%	-61%
Output share of small firms	Ω_1	0.136	0.035	-73.9%	-73%
Panel B: Markups and Misallocation					
Life cycle of markups		8.2%	6.2%	0.02	-
Average markup	$E[\mu]$	11.73%	9.25%	-0.025	-
Dispersion in log markups across products	$\sigma(\ln \mu)$	10.5%	8.47%	-19.3%	-
Dispersion in log markups across firms	$\sigma(\ln \mu_f)$	8.1%	5.86%	-27.5%	-
Aggregate misallocation	$\mathcal{M}$	0.995	0.997	0.2%	-
labor wedge	Λ	0.887	0.903	1.8%	-
Panel C: Innovation, Expansion and Entry and Growth					
Rate of growth	g	3%	2.91%	-0.0009	-0.0004
Own innovation	I	0.561	0.498	-11.1%	-8.3%
Incumbent creative destruction	x	0.207	0.267	28.6%	23.4%
Aggregate creative destruction	τ	0.240	0.277	15.3%	12.8%

Notes: The first panel contains the calibration moments. The entry rate is simply the share of firms, which are entrants. For the US, the employment life cycle is calculated as average employment of firms between 10 and 14 years old relative to firms with age less than 5 and stems from Hsieh and Klenow (2014). The parameters for the Indonesian economy are contained in Table 3. For the US economy, I recalibrate the relative efficiency of expansion and entry, i.e. φ_z and φ_x. All remaining parameters are the same as in Table 3. In the last column I report the results of a calibration of a model with constant markups. See Section OA-1.6 in the Online Appendix for details.

Table 6: A Counterfactual: Changes in Expansion and Entry Costs

(Panel B) and the sources of aggregate growth (Panel C). Panel A shows that the estimated reductions in entry and expansion costs affect the firm size distributions markedly. As firms have more opportunities to expand their scale of production, the firm size distribution shifts to the right and the economy sustains less firms in equilibrium. The equilibrium number of firms declines by 60%, i.e. average firm size more than doubles. This reallocation comes especially at the expense of small firms so that the sales share of firms that only produce a single product, declines by more than 70%. Hence, seemingly small differences in the rate of entry and life cycle growth have large effects on the cross-sectional distribution of firm size.

Panel B shows - as implied by Proposition 4 - that this shift towards large firms is accompanied with pro-competitive effects. Even though markups are increasing in size in the cross-section, lower entry and expansion costs increase average firm size and simultaneously reduce markups as firms increase their markups at a lower rate. In particular, while seven year old firms in Indonesia have about 8% higher markups than current entrants, this difference declines by two percentage points in the US.[30] This reduction in markup growth reduces average markups and their dispersion. Both the welfare-relevant dispersion of markups at the product level and the empirically measured dispersion of markups at the firm level decline by 20-30%. Note in particular, that the firm-level dispersion declines more: as firms become larger, the measured markup dispersion is less informative about the actual dispersion of product level prices. The last two columns show that this change in the distribution of markups lowers misallocation by about one-third. In particular, TFP increases by 0.2% and the reduction in monopoly power is akin to a 1.8% decline in taxes on static factors.

Finally, Panel C consider the implications on aggregate growth. The most striking result is that the economy-wide growth rate g hardly changes - if anything it slightly declines once entry and expansion barriers are dismantled. The reason is the equilibrium effect on firms' own-innovation incentives. While creative destruction increases by 15%, firms' incentives to increase productivity within their existing markets decline by 11%. That these different margins of growth are negatively related is not surprising (recall the optimality condition in (10), which showed that the marginal value to accumulate markups is discounted at rate $\rho + \tau$). This competition effect is sufficiently strong that $I + \tau$ declines, even though τ increases.

The results in Table 6 have four implications. First of all, seemingly large changes in the stationary firm size distribution and the number of active firms are fully consistent with empirically plausible small differences in observable entry rates, employment life cycle growth and the increase in markups by age. Secondly, such large differences do not imply that countries are predicted to grow at vastly different rates. A growth differences of the one reported in Table 6 only accumulates to a productivity level difference of 2% after 20 years. Third, the results suggest that frictions for existing firms to expand into new markets are more important than differences in entry costs to understand the empirical firm-level patterns across countries. While such frictions readily imply that many firms are small and experience little growth as they age, high entry costs would have the exact opposite implication. Finally, while misallocation is indeed predicted to be lower, the quantitative magnitude is relatively small, as the static efficiency losses of markups are limited to begin with. To the extent that misallocation is an important determinant of cross-country differences in aggregate TFP, it is unlikely that differences in markup heterogeneity account for a substantial share of it.

To highlight which of these implications are due to the presence of heterogeneous markups per se, in the last column of Table 6 I report the results of a calibration of a model with constant markups. This model, which I characterize in detail in Section OA-1.6 in the Online Appendix, has the same structure as the model with CES preferences laid out in Section 2.9 except that I assume that firms always charge a constant markup μ. Because

[30] This is qualitatively consistent with the results reported in Hsieh and Klenow (2014), who argue that the increase in revenue-productivity by age, which in my model is proportional to markups, is steeper in India relative to the US.

profits are increasing in quality firms have an incentive to engage in own-innovation even if markups are constant. In order to compare this model to the model with heterogenous markups I calibrate it to the same moments. I pick the exogenous markup μ to generate the same average markup as in the model with heterogeneous markups. Given the life-cycle profile of markups is by construction flat, I need to introduce one additional moment to separately identify the cost of own-innovation φ_I from the cost of expansion φ_x. To do so, I assume that own-innovation accounts for the same share of growth as in the model with heterogenous markups (i.e. 70%). While the model has of course no implications for the pattern of markups, Table 6 shows that it also implies that frictions for existing firms are more important than entry costs, that firms are substantially smaller in the US and that the effects of the higher rate of creative destruction on the equilibrium growth rate are small once the endogenous response of firms' own-innovation activities are taken into account.

Robustness In Section OA-2.6 in the Online Appendix I examine the robustness of these results. First of all, I show that the results do not depend substantially on the choice of the curvature parameter ζ. The resulting changes in the number of firms or the importance of small firms are all very similar to the ones reported in Table 6 and the quantitative effect on the equilibrium growth rate is very small (even though for large values of ζ, the predicted growth rate in the US is slightly higher). I also study the sensitivity with respect to the underlying moments, i.e. the extent of life cycle growth and the extent of entry. The elasticity of average firm size and and the share of small firms with respect to these two moments is quite sizable. If one were to for example assume that the extent of life cycle growth in the US was 2.5 instead of 2, the number of firms and the share of small firms would fall by 90%. In contrast, the effect on the growth rate is still almost indistinguishable from zero. In Section OA-2.8 in the Online Appendix I also report an alternative calibration, which does not rely on the observed entry rate but explicitly exploits the size cutoff of census data.

3.5 Expansion and Entry Costs in Indonesia: Empirical Evidence

The analysis above suggests that differences in expansion costs could be an important determinant of firm size, firm growth and misallocation. In this last section I provide some suggestive evidence for such frictions. To do so I exploit regional variation across products markets in Indonesia. As I do not have direct information on the type of barriers different firms might face, I use the theory to suggest an empirical strategy based on the joint patterns of various firm-level outcomes.

The basic intuition is simple. If different regions in Indonesia differed only in their expansion costs, locations with low frictions should see fewer and bigger firms, lower entry rates and a steeper schedule of life cycle employment growth. Additionally, product markets in such regions should be characterized by lower markups. If in contrast entry costs were the dominant source of variation, it would also be the case that firm size should be negatively correlated with regional entry rates and positively correlated with the slope of life cycle growth - however, the underlying source of variation would be exactly reversed. Now large firms should reside in regions with high entry costs and one would expect a positive correlation between firm size and the prevailing markups.

I implement this strategy in the following way. The Indonesian micro-data allows me to link individual firms to their geographic location. I define a geographical region as a province, of which there are 27 in the data. Because I do not have information on where firms sell their products, I need to assume that firms are predominantly active in their own province. Provinces obviously differ in their industrial composition. As industries differ in their average size, I conduct the entire analysis at the region-industry level and control for the common industry component using fixed effects. Hence, the variation of interest is geographical in nature. More specifically, I calculate my outcomes of interest, i.e. average firm size, entry and exit rates, average markups and the employment life cycle growth rates

	Entry rate	Exit rate	LC empl growth	Markups Avg.	q^{90}
Average firm size	-0.019***	-0.011***	0.132***	-0.054***	-0.049***
	(0.004)	(0.003)	(0.034)	(0.014)	(0.010)
Controls	Local population; Local agricultural share; Industry FE; Year FE				
N	455	463	463	462	462
R^2	0.369	0.320	0.640	0.378	0.369

Notes: Standard errors are clustered at the level of a province and contained in parentheses with ***, **, and * respectively denoting significance at the 1%, 5% and 10% levels. Regression are run at the province-industry level, where industries are measured at the 3-digit level. The variables are all measured within these province-industry cells. The entry and exit rates are measured as the share of entering and exiting firms. The employment life cycle is measured as the growth of cohort employment over the 3 year horizon. Log markups are measured as the residual from a regression of log inverse labor shares on a set of year and 5-digit industry fixed effects and "Avg" is the mean log markup and q^{90} is the 90% quantile. "Average firm size" is the average log value added within a industry-region-year cell. All regressions contain a full set of industry and year fixed effects and control for the log of the province population and the share of villages within the province, which are agricultural. I only consider province-industry cells with at least 50 observations and all regressions are weighted using the number of observations within each cell as a weights.

Table 7: Firm size, Entry and Markups across Product Markets in Indonesia

for each province-industry-year cell and then consider regressions of the form

$$y_{rst} = \delta_s + \delta_t + \beta \times AvgSize_{rst} + \gamma \times \ln(pop_r) + \alpha \times Ag_r + u_{rst},$$

where δ_s and δ_t are industry and time fixed effects, $AvgSize_{rst}$ is the average size of producers active in region r, in sector s in time t and y_{rst} are the different outcome variables mentioned above. Moreover, I also control for the size of the population in region r and the regional agricultural share to account for the effects of market size.[31] Given the focus on the regional variation, I cluster all standard errors at the province level, to allow for correlation in the error term across industries within a province.

My preferred measure of size is firms sales (rather then employment), as the theory predicts that average sales only depend on $\vartheta_x = x/\tau$. To calculate the employment life cycle, I again rely on the panel dimension and adopt the same methodology as for Figure 4.[32] To not identify the parameters from sparsely populated region-industry-year cells, I only consider observations, which contain at least 50 observations and I weight the regression by the number of observations in each bin.

Table 7 contains the results. In the first three columns I show that average firm size in a region is negatively correlated with entry and exit rates and positively correlated with the extent of life cycle growth. These correlations hold regardless of whether the source of variation across regions stems from entry or expansion costs. Columns four and five show that average size is negatively correlated with markups as measured by either the average markup or the 90%-quantile. This is consistent with the model if regional firm size is driven by differences in expansion costs but not consistent with an explanation based on entry costs. While these patterns are consistent with expansion costs potentially playing an important role in the determination of misallocation, firm dynamics and growth, they are of course only suggestive. They however highlight the need to directly measure why the costs of entering products markets within countries might be systematically related to the level of development and whether these are amenable by policy.[33]

[31] To measure the geographical characteristics, I exploit information from the Village Potential Statistics (PODES) dataset in 1996. The PODES dataset contains detailed information on all of Indonesia's 65.000 villages. Using the village level data, I then aggregate this information to the province level and match these to the firm-level data. In particular I measure the size of the population and the share of the population living in villages within a region, which are predominantly agricultural.

[32] In particular, for each cohort I calculate employment growth at the three year horizon after controlling for a full set of 5-digit product and year fixed effects and than average these growth rates at the industry-province level. I have to opt for a somewhat short horizon of three years to not lose too many observations given that I only have data for the years 1991 to 1998.

[33] In section OA-2.6 in the Appendix I show that the results are robust to (a) measuring firms size by employment, (b) not weighting the regressions, (c) different choices to the minimum number of firms for a given cell to appear in the final sample, (d) not clustering the standard errors and (e) different measures of entry.

4 Conclusion

This paper proposes a novel model of firm-dynamics where firms' market power is endogenous and the distribution of markups emerges as an equilibrium outcome. The theory is highly tractable, can be solved analytically and provides a unifying framework to link firm-growth, markups, misallocation and aggregate growth.

The central economic idea of this paper is that markups emerge as the solution of a forward-looking, risky accumulation problem. Firms invest resources to increase the productivity of their existing products. Doing so allows them to rack up markups by pulling away from competing firms. At the same time they are subject to the threat of creative destruction, whereby more efficient competitors enter and markups decline as competition intensifies. The extent of churning therefore keeps monopoly power in check and emerges as the key determinant of the equilibrium distribution of markups and the macroeconomic costs of misallocation.

Despite its parsimony, the theory has rich empirical predictions for the relationships between markups, size and age at the firm-level. Firms own a portfolio of products and markups follow a distinct life cycle pattern. In particular, firms grow in size by expanding into new, low-markup products and increase their profitability by slowly accumulating market power in the products they own. I show that this implies that markups are increasing in age for the majority of firms and that lower creative destruction increases the extent of markup life cycle growth and raises markups, especially in the tail of the markup distribution.

As an application, I calibrate the model to firm-level panel data from Indonesia. This setting is motivated by the recent literature on misallocation in developing countries. This literature typically takes the extent of misallocation as given and relies on exogenous firm-specific wedges as a modeling device. In the model presented in this paper, such wedges exactly coincide with firms' markups and are therefore endogenously determined. I find that markups can plausibly account for 15% of the measured dispersion in average products and could reduce aggregate TFP by roughly 1% if they were the sole source of misallocation. The static efficiency losses from markups alone are therefore unlikely to be main culprit of aggregate TFP differences across countries. I also find that firms in Indonesia are particularly hurt by frictions for existing firms to enter new product markets. In contrast, higher entry costs for new firms are less important. Quantitatively, such frictions increase of misallocation by 0.3% and reduce average firm size substantially, without markedly affecting the endogenous growth rate. Large differences in the distribution of firm size are therefore consistent with a stable distribution of income across countries.

References

Acemoglu, Daron and Ufuk Akcigit, "Intellectual Property Rights Policy, Competition and Innovation," *Journal of the European Economic Association*, 2012, *10* (1), 1–42.

_ , _ , **Nicholas Bloom, and William Kerr**, "Innovation, Reallocation and Growth," 2016. Working Paper.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

_ , **Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li**, "A theory of falling growth and rising rents," 2019.

_ , **Christopher Harris, Peter Howitt, and John Vickers**, "Competiton, Imitation and Growth with Step-By-Step Innovation," *Review of Economic Studies*, 2001, *68*, 467–492.

_ , **Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt**, "Competition and Innovation: An Inverted U-Relationship," *Quarterly Journal of Economics*, 2005, pp. 701–26.

_ , **Robin Burgess, Steve Redding, and Fabrizio Zilibotti**, "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India," *The American Economic Review*, 2008, *94* (4), 1397–1412.

Akcigit, Ufuk and Sina T Ates, "What Happened to US Business Dynamism?," Technical Report, National Bureau of Economic Research 2019.

_ **and William Kerr**, "Growth through Heterogeneous Innovations," *Journal of Political Economy*, 2018, *126* (4), 1374–1443. Working Paper.

_ , **Harun Alp, and Michael Peters**, "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," 2015. Working Paper.

Amiti, Mary and Jozef Konings, "Trade Liberalization, Intermediate Inputs and Productivity," *American Economic Review*, 2007, *97* (5), 1611–1638.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare, "The Elusive Pro-Competitive Effects of Trade," 2016. Working Paper.

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

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

_ **and** _ , "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 2010, *118* (3), 433–484.

_ **and** _ , "Aggregate Implications of Innovation Policy," 2015. Working Paper.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, John Van Reenen et al., *The fall of the labor share and the rise of superstar firms*, National Bureau of Economic Research, 2017.

Barseghyan, Levon, "Entry Costs and Cross-Country Differences in Productivity and Output," *Journal of Economic Growth*, 2008, *13* (2), 145–167.

_ **and Riccardo DiCecio**, "Entry Costs, Misallocation and Cross-Country Income and TFP Differences," Working Paper, Federal Reserve Bank of St. Louis 2009.

Bartelsman, Eric J., John C. Haltiwanger, and Stefano Scarpetta, "Cross-Country Differences in Productivity: The Role of Allocative Efficiency," *American Economic Review*, 2013, *103* (1), 305–334.

Bento, Pedro, "Competition, Innovation, and the Number of Firms," 2016. Working Paper.

_ **and Diego Restuccia**, "Misallocation, Establishment Size, and Productivity," *American Economic Journal: Macroeconomics*, 2017, *9* (3), 267–303.

Bernard, Andrew B., Stephen Redding, and Peter Schott, "Multi-Product Firms and Product Switching," *American Economic Review*, 2010, *100* (1), 70–97.

Bernard, Andrew, Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Blalock, Garrick, Paul J. Gertler, and David I. Levine, "Financial Constraints on Investment in an Emerging Market Crisis," *Journal of Monetary Economics*, 2008, *55* (3), 568–591.

Blas, Beatriz De and Katheryn N Russ, "Understanding markups in the open economy," *American Economic Journal: Macroeconomics*, 2015, *7* (2), 157–80.

Buera, Francisco and Roberto N. Fattal Jaef, "The Dynamics of Development: Innovation, and Reallocation," 2016. Working Paper.

_ , Joe Kaboski, and Yongseok Shin, "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 2011, *101* (5), 1964–2002.

Clementi, Gian Luca and Hugo Hopenhayn, "A Theory of Financing Constraints and Firm Dynamics," *The Quarterly Journal of Economics*, 2006, *121* (1), 229–265.

Da-Rocha, José-María, Marina Mendes Tavares, and Diego Restuccia, "Policy distortions and aggregate productivity with endogenous establishment-level productivity," Technical Report, National Bureau of Economic Research 2017.

David, Joel M and Venky Venkateswaran, "The sources of capital misallocation," *American Economic Review*, 2019, *109* (7), 2531–67.

_ , Hugo A Hopenhayn, and Venky Venkateswaran, "Information, misallocation, and aggregate productivity," *The Quarterly Journal of Economics*, 2016, *131* (2), 943–1005.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "The role of entrepreneurship in US job creation and economic dynamism," *Journal of Economic Perspectives*, 2014, *28* (3), 3–24.

Dhyne, Emmanuel, Amil Petrin, Valerie Smeets, and Frederic Warzynski, "Multi product firms, import competition, and the evolution of firm-product technical efficiencies," Technical Report, National Bureau of Economic Research 2017.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-De-Silvaes, and Andrei Shleifer, "The Regulation of Entry," *Quarterly Journal of Economics*, 2002, *117* (1), 1–37.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups and the Gains from International Trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

_ , _ , and _ , "How Costly Are Markups?," 2018. Working Paper.

Epifani, Paolo and Gino Gancia, "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 2011, *83* (1), 1–13.

Fattal Jaef, Roberto N., "Entry and Exit, Multi-Product Firms, and Allocative Distortions," *American Economic Journal: Macroeconomics*, 2018, *10* (2), 86–112. Working Paper.

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

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 1999, *114* (3), 739–767.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Pete Klenow, "How Destructive is Innovation?," 2016. Working Paper.

Grossman, Gene and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 1991, *58* (1), 43–61.

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

Haltiwanger, John, Ryan Decker, and Ron Jarmin, "Top ten signs of declining business dynamism and entrepreneurship in the US," in "Kauffman Foundation New Entrepreneurial Growth Conference" Citeseer 2015.

Herrendorf, Berthold and Arilton Teixeira, "Barriers to Entry and Development," *International Economic Review*, 2011, *52* (2), 573–602.

Holmes, Thomas J., Wen-Tai Hsu, and Sanghoon Lee, "Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade," *Journal of International Economics*, 2014, *92* (2), 195–206.

Hopenhayn, Hugo, "On the Measure of Distortions," *Working Paper*, 2012.

Hsieh, Chang-Tai and Benjamin A. Olken, "The Missing Missing Middle," *Journal of Economic Perspectives*, 2014.

_ **and Peter Klenow**, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (2), 771–807.

_ **and** _ , "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 2014, *129* (3), 1035–1084.

Hurst, Erik and Ben Pugsley, "What Do Small Businesses Do," *Brookings Papers on Economic Activity*, 2012.

Jones, Charles and Jihee Kim, "A Schumpeterian Model of Top Income Inequality," 2016. Working Paper.

Karahan, Fatih, Benjamin Pugsley, and Aysegül Sahin, "Understanding the 30-year Decline in the Start-up Rate: A General Equilibrium Approach," 2015. Working Paper.

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

Lentz, Rasmus and Dale T. Mortensen, "An Empirical Model of Growth through Product Innovation," *Econometrica*, 2008, *76* (6), 1317–1373.

Loecker, Jan De, "Recovering Markups from Production Data," *International Journal of Industrial Organization*, 2011, *29*, 350–355.

_ **and Frederic Warzynski**, "Markups and Firm-Level Export Status," *American Economic Review*, 2012, *102* (6), 2437–2471.

_ **and Jan Eeckhout**, "The rise of market power and the macroeconomic implications," Technical Report, National Bureau of Economic Research 2017.

_ **, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik**, "Prices, Markups and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Luttmer, Erzo G. J., "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, 2010, *2*, 547–576.

Melitz, Marc and Giancarlo Ottaviano, "Market Size, Trade and Productivity," *Review of Economic Studies*, 2008, *75* (1), 295–316.

Midrigan, Virgiliu and Daniel Yi Xu, "Finance and Misallocation: Evidence from Plant-Level Data," *The American Economic Review*, 2014, *104* (2), 422–458.

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

Peters, Michael and Conor Walsh, "Declining Dynamism, Increasing Markups and Missing Growth: The Role of The Labor Force," 2019. Working Paper.

Porta, Rafael La and Andrei Shleifer, "The Unofficial Economy and Economic Development," in "Brookings Papers on Economic Activity" 2009, pp. 275–352.

_ **and** _ , "Informality and Development," *Journal of Economic Perspectives*, 2014, pp. 109–126.

Restuccia, Diego and Richard Rogerson, "Policy Distortions and Aggregate Productivity with Heterogenous Establishments," *Review of Economic Dynamics*, 2008, *11* (4), 707–720.

Schoar, Antoinette, "The Divide between Subsistence and Transformational Entrepreneurship," in "Innovation Policy and the Economy, Volume 10," University of Chicago Press, 2010, pp. 57–81.

Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, 1997, *35*, 40–59.

Yang, Mu-Jeung, "Micro-level Misallocation and Selection: Estimation and Aggregate Implications," *Working Paper*, 2012.

Appendix

A-1 Appendix: Theoretical Results

A-1.1 Proof of Proposition 1

I prove Proposition 1 in two steps. I first derive the value function (8). Then I show that the conditions in Proposition 1 uniquely define the optimal choices for (I, x, z).

A-1.1.1 Solving for the value function $V_t(.)$ in (8)

To derive (8), conjecture first that the value function takes an additive form[34]

$$V_t\left(n, [\Delta_i]_{i=1}^n\right) = V_t^P(n) + \sum_{i=1}^n V_t^M(\Delta_i). \tag{A-1}$$

V_t^P and V_t^M are therefore defined by the differential equations

$$rV_t^M(\Delta) - \dot{V}_t^M(\Delta) = \pi_t(\Delta) - \pi_t(1) - \tau V_t^M(\Delta) + \max_I \left\{ I\left[V_t^M(\Delta+1) - V_t^M(\Delta)\right] - c^I(I,\Delta)\, w_t \right\}, \tag{A-2}$$

and

$$\begin{aligned} rV_t^P(n) - \dot{V}_t^P(n) &= n \times \pi_t(1) + \sum_{i=1}^n \tau\left[V_t^P(n-1) - V_t^P(n)\right] \\ &\quad + \max_X \left\{ X\left[V_t^P(n+1) + V_t^M(1) - V_t^P(n)\right] - c^X(X,n)\, w_t \right\}. \end{aligned}$$

Now consider a steady state where both value functions grow at rate g and assume the cost functions in (7). Then we can write (A-2) as

$$(r+\tau-g)\, V_t^M(\Delta) = \pi_t(\Delta) - \pi_t(1) + \max_I \left\{ I\left[V_t^M(\Delta+1) - V_t^M(\Delta)\right] - \frac{1}{\varphi_I}\lambda^{-\Delta} I^{\zeta} w_t \right\}.$$

Conjecture that

$$V_t^M(\Delta) = \kappa_t - \alpha_t \lambda^{-\Delta}. \tag{A-3}$$

Then $V_t^M(\Delta+1) - V_t^M(\Delta) = \frac{\lambda-1}{\lambda}\alpha_t\lambda^{-\Delta}$. This implies that optimal innovation rate I solves

$$I_t = \left(\frac{\lambda-1}{\lambda}\frac{\varphi_I}{\zeta}\frac{\alpha_t}{w_t}\right)^{\frac{1}{\zeta-1}}. \tag{A-4}$$

Suppose that $\frac{\alpha_t}{w_t}$ is constant along the BGP (which I will verify below). (A-4) then implies that $I_t(\Delta) = I$. Using (A-4), (A-3) and the Euler equation $\rho = r - g$ yields

$$(\rho+\tau)\left[\kappa_t - \alpha_t\lambda^{-\Delta}\right] = \left(\frac{1}{\lambda} - \lambda^{-\Delta}\right) Y_t + \frac{\zeta-1}{\varphi_I}\lambda^{-\Delta} I^{\zeta} w_t,$$

[34] The analysis in this section only contains the most important steps. A detailed derivation is contained in Section OA-1.1 of the Online Appendix.

so that $\kappa_t = \frac{\frac{1}{\lambda} Y_t}{\rho+\tau}$ and $\alpha_t = \frac{Y_t - \frac{\zeta-1}{\varphi_I} I^\zeta w_t}{\rho+\tau}$. Hence, (A-3) yields

$$V_t^M(\Delta) = \frac{\left(\frac{1}{\lambda} - \lambda^{-\Delta}\right) Y_t + \lambda^{-\Delta} \frac{\zeta-1}{\varphi_I} I^\zeta w_t}{\rho+\tau} = \frac{\pi_t(\Delta) - \pi_t(1) + (\zeta-1) c_I(I,\Delta) w_t}{\rho+\tau}.$$

Note also that this implies that

$$V_t^M(1) = \frac{(\zeta-1) w_t \frac{1}{\lambda} \frac{1}{\varphi_I} I^\zeta}{\rho+\tau}. \tag{A-5}$$

Now turn to $V_t^P(n)$. Define $X = xn$ and conjecture that $V_t^P(n) = n \times v_t$, where v_t grows at rate $g = r - \rho$. Hence,

$$(\rho+\tau) v_t = \pi_t(1) + \max_x \left\{ x \left[v_t + V_t^M(1) \right] - \frac{1}{\varphi_x} x^\zeta w_t \right\}$$

The optimality condition for x reads

$$v_t + V_t^M(1) = \frac{\zeta}{\varphi_x} x^{\zeta-1} w_t. \tag{A-6}$$

As v_t and $V_t^M(\Delta)$ both grow at rate g, this implies that x is indeed constant. In particular, given x, v_t is given by

$$(\rho+\tau) v_t = \pi_t(1) + (\zeta-1) \frac{1}{\varphi_x} x^\zeta w_t. \tag{A-7}$$

To solve for v_t, let $v_t = \overline{v} \times w_t$. The unknowns $(x, \overline{v})$ are then determined from (A-6) ,(A-7) and (A-5) as

$$\begin{aligned} \frac{\zeta}{\varphi_x \phi_e^\zeta} x^{\zeta-1} &= \overline{v} + \frac{V_t^M(1)}{w_t} = \overline{v} + \frac{(\zeta-1) \frac{1}{\lambda} \frac{1}{\varphi_I} I^\zeta}{\rho+\tau} \\ \overline{v} &= \frac{\frac{\lambda-1}{\lambda} \frac{Y_t}{w_t} + \frac{\zeta-1}{\varphi_x} x^\zeta}{\rho+\tau}. \end{aligned}$$

The final value function $V_t^P(n)$ is given by

$$V_t^P(n) = n \times \frac{\pi_t(1) + (\zeta-1) c^X(1,x) w_t}{\rho+\tau}.$$

Substituting into (A-1) yields (8).

A-1.1.2 Existence and Uniqueness

We now prove existence and uniqueness of the equilibrium. We need to solve for the tupel (I, x, z). Alternatively, we can solve for (I, x, τ) and then solve for $z = \tau - x$. From the static allocations we know that $Y_t \Lambda_t = w_t L_t^P$. Note that Λ_t is a known function of τ/I (see Proposition 2) and hence I write it as $\Lambda\left(\frac{\tau}{I}\right)$. To solve for L_t^P we need the labor market clearing condition, which is given by $1 = L_t^P + L_t^I + L_t^x + L_t^z$,where L_t^j denotes the total amount of labor used for the respective sources of innovation, expansion and entry. Note that $L_t^z = \frac{1}{\varphi_I} z$, $L_t^x = \frac{1}{\varphi_x} x^\zeta$ and

$$L_t^I = \int_{j=0}^1 c^I(I, \Delta_j)\, dj = \int_{j=0}^1 \frac{1}{\varphi_I} \times I^\zeta \lambda^{-\Delta_j} dj = \frac{1}{\varphi_I} \times I^\zeta \times \Lambda\left(\frac{\tau}{I}\right).$$

Hence, the equilibrium is defined by the four equations

$$
\begin{aligned}
1 &= \Lambda\left(\frac{\tau}{I}\right) \times \frac{Y_t}{w_t} + \frac{1}{\varphi_I} \times I^{\zeta} \times \Lambda\left(\frac{\tau}{I}\right) + \frac{1}{\varphi_I} \times (\tau - x) + \frac{1}{\varphi_x} \times x^{\zeta} && \text{(A-8)}\\
\frac{1}{\varphi_z} &= \frac{\frac{\lambda-1}{\lambda}\frac{Y_t}{w_t} + \frac{\zeta-1}{\varphi_x}x^{\zeta} + \frac{\zeta-1}{\varphi_I}\frac{1}{\lambda}I^{\zeta}}{\rho+\tau} && \text{(A-9)}\\
\frac{Y_t}{w_t} &= \frac{\lambda}{\lambda-1}\frac{\zeta}{\varphi_I}I^{\zeta-1}(\rho+\tau) + (\zeta-1)\frac{1}{\varphi_I}I^{\zeta} && \text{(A-10)}\\
\frac{Y_t}{w_t} &= \frac{\zeta}{\varphi_x}x^{\zeta-1}(\rho+\tau)\frac{\lambda}{\lambda-1} - \frac{\lambda}{\lambda-1}\frac{\zeta-1}{\varphi_x}x^{\zeta} - \frac{1}{\lambda-1}\frac{\zeta-1}{\varphi_I}I^{\zeta}. && \text{(A-11)}
\end{aligned}
$$

To solve for the unknowns $\left(\frac{Y}{w}, I, \tau, x\right)$, note first that (A-11) and (A-9) imply that

$$
x^{\zeta-1} = \frac{\varphi_x}{\varphi_z}\frac{1}{\zeta}. \tag{A-12}
$$

This determines x in terms of parameters. We can then use (A-9), (A-10) and (A-8) to arrive at two equations in the two unknowns (τ, I)

$$
\begin{aligned}
1 &= \Lambda\left(\frac{\tau}{I}\right) \times \left(\frac{\lambda}{\lambda-1}\frac{\zeta}{\varphi_I}I^{\zeta-1}(\rho+\tau) + \zeta\frac{1}{\varphi_I}I^{\zeta}\right) + \frac{1}{\varphi_z} \times \tau - h(\varphi). && \text{(A-13)}\\
\frac{1}{\varphi_z} &= \frac{\zeta}{\varphi_I}I^{\zeta-1} + \frac{\frac{\zeta-1}{\varphi_I}I^{\zeta}}{\rho+\tau} + \frac{h(\varphi)}{\rho+\tau}, && \text{(A-14)}
\end{aligned}
$$

where

$$
h(\varphi) = \left(\frac{\zeta-1}{\zeta}\right)\left(\frac{\varphi_x}{\zeta\varphi_z^{\zeta}}\right)^{\frac{1}{\zeta-1}} \geq 0. \tag{A-15}
$$

Given a solution (I, τ) and x from (A-12), we can calculate $\frac{Y}{w}$ from (A-10) and $z = \tau - x$. Hence, we only have to show that (A-13) and (A-14) have a unique solution. Rewriting (A-13) and (A-14) in terms of $\vartheta_I = s$ yields

$$
\begin{aligned}
1 &= \Lambda(s) \times \left(\frac{\lambda}{\lambda-1}\frac{\zeta}{\varphi_I}I^{\zeta-1}(\rho+sI) + \zeta\frac{1}{\varphi_I}I^{\zeta}\right) + \frac{1}{\varphi_z} \times sI - h(\varphi) && \text{(A-16)}\\
\frac{1}{\varphi_z} &= \frac{\zeta}{\varphi_I}I^{\zeta-1} + \frac{\frac{\zeta-1}{\varphi_I}I^{\zeta}}{\rho+sI} + \frac{h(\varphi)}{\rho+sI}, && \text{(A-17)}
\end{aligned}
$$

where $\Lambda'(s) > 0$. Let us write the first equation as

$$
1 = H(I, s) \equiv \Lambda(s) \times \left(\frac{\lambda}{\lambda-1}\frac{\zeta}{\varphi_I}I^{\zeta-1}(\rho+sI) + (\zeta-1)\frac{1}{\varphi_I}I^{\zeta}\right) + \frac{1}{\varphi_I} \times I^{\zeta} \times \Lambda(s) + \frac{1}{\varphi_z} \times sI - h(\varphi).
$$

Then $\frac{\partial H(I,s)}{\partial s} > 0$ and $\frac{\partial H(I,s)}{\partial I} > 0$. Hence, (A-16) defines a downward sloping continuous schedule in the (s, I) space, which we call $I^{LM}(s)$. Moreover, $\lim_{s\to\infty} I^{LM}(s) \to 0$ and $\lim_{s\to 0} I^{LM}(s) \to \infty$. Now write the second equation, which stems from the free entry condition as

$$
\frac{1}{\varphi_z} = \frac{\zeta}{\varphi_I}I^{\zeta-1} + \frac{1}{\rho+sI}\left(\frac{\zeta-1}{\varphi_I}I^{\zeta} + h(\varphi)\right) = G\left(I^{FE}(s), s\right). \tag{A-18}
$$

Clearly, $\frac{\partial G(I,s)}{\partial s} < 0$. Also

$$\begin{aligned}\frac{\partial G}{\partial I} &= \frac{(\zeta-1)\zeta}{\varphi_I} I^{\zeta-1}\frac{1}{I} - (\rho+sI)^{-2} s\left[(\zeta-1)\frac{1}{\varphi_I}I^{\zeta}+h(\varphi)\right] + (\rho+sI)^{-1}(\zeta-1)\zeta\frac{1}{\varphi_I}I^{\zeta-1} \\ &= (\rho+sI)^{-2}\left[\frac{(\zeta-1)\zeta}{\varphi_I}I^{\zeta-1}\frac{(\rho+sI)^2}{I} - s\left[(\zeta-1)\frac{1}{\varphi_I}I^{\zeta}+h(\varphi)\right] + (\rho+sI)(\zeta-1)\zeta\frac{1}{\varphi_I}I^{\zeta-1}\right].\end{aligned}$$

Given the definition of h in (A-15), it can be shown that $\frac{\partial G(I,s)}{\partial I} > 0$. Hence, $\frac{\partial I^{FE}}{\partial s} > 0$. Note that $I^{FE}(s)$ has to be bounded for (A-18) to be satisfied. Hence, $0 \leq I^{FE}(s) \leq I^{max}$. This also implies that $\lim_{s\to\infty} sI^{FE}(s) = \infty$, so that $\lim_{s\to\infty} I^{FE}(s) \to \left(\frac{\varphi_I}{\varphi_z}\frac{1}{\zeta}\right)^{\frac{1}{\zeta-1}} = I^{max}$. Now consider the case of $s \to 0$. As $I^{FE}(s)$ is declining in s it has to be that $sI^{FE}(s) \to 0$. Let $I^{FE}(0)$ be that limit. (A-18) then implies that $I^{FE}(0)$ is implicitly defined by $\frac{1}{\varphi_z} - \frac{h(\varphi)}{\rho} = \frac{\zeta}{\varphi_I} I^{FE}(0)^{\zeta-1} + \frac{\frac{\zeta-1}{\varphi_I} I^{FE}(0)^{\zeta}}{\rho}$. As $I^{FE}(0) > 0$, this requires that $\frac{1}{\varphi_z} > \frac{h(\varphi)}{\rho}$. From (A-15) we can write this condition as

$$\rho > \varphi_z\left(\frac{\zeta-1}{\zeta}\right)\left(\frac{\varphi_x}{\zeta\varphi_z^{\zeta}}\right)^{\frac{1}{\zeta-1}} = \left(\frac{\zeta-1}{\zeta}\right)\left(\frac{\varphi_x}{\zeta\varphi_z}\right)^{\frac{1}{\zeta-1}}. \tag{A-19}$$

As long as (A-19) is satisfied, there is a unique solution (I,s) for the system of equations (A-16) and (A-17). Hence, there is a unique $\tau = s \times I$. The optimal expansion rate x is given by (A-12).

A-1.2 Proof of Proposition 2

Consider first the distribution of quality gaps $\nu(\Delta,t)$. In a stationary equilibrium we have $\dot{\nu}(\Delta,t) = 0$. (11) then implies that

$$\nu(\Delta) = \left(\frac{I}{\tau+I}\right)^{\Delta}\frac{\tau}{I} = \left(\frac{1}{1+\frac{\tau}{I}}\right)^{\Delta}\frac{\tau}{I} = \left(\frac{1}{1+\vartheta_I}\right)^{\Delta}\vartheta_I$$

Hence, $P(\Delta \leq d) = 1 - \left(\frac{1}{1+\vartheta_I}\right)^d = 1 - e^{-\ln(1+\vartheta_I)\times d}$. This implies that log markups $\ln(\mu) = \Delta\ln(\lambda)$ are exponentially distributed with parameter θ. Similarly, $F(\mu;x) = P\left(\lambda^{\Delta} \leq \mu\right) = 1 - \mu^{-\theta}$. To derive (12), note that[35]

$$\begin{aligned}\Lambda &= \int \mu^{-1}\theta\mu^{-(\theta+1)}d\mu = \frac{\theta}{\theta+1} \\ M &= exp(-E\ln(\mu))\Lambda^{-1} = e^{-\theta^{-1}}\frac{\theta+1}{\theta}.\end{aligned}$$

A-1.3 Proof of Proposition 3

In this section I derive the life cycle properties of markups.

The distribution of markups as a function of product age: equation (13) I first show that the distribution of quality gaps Δ as a function of age conditional on survival, $\zeta_{\Delta}(a)$, is given by $\zeta_{\Delta+1}(a) = \frac{1}{\Delta!}(Ia)^{\Delta}e^{-Ia}$. Let $p_{\Delta}(a)$ denote the probability of the product having a quality gap Δ at age a when it was introduced at time 0. The corresponding flow

[35] Note that these expression are derived taking the distribution of markups as continuous, even though the model implies that they are discrete. Taking this discreteness explicitly into account yields

$$\Lambda^{Dis} = \sum_{i=1}^{\infty}\lambda^{-i}\mu(i) = \frac{\frac{\tau}{I}}{\lambda\left(\frac{\tau}{I}+1\right)}\sum_{i=0}^{\infty}\left(\frac{1}{\lambda\left(\frac{\tau}{I}+1\right)}\right)^i = \frac{\frac{\tau}{I}}{\lambda\frac{\tau}{I}+\lambda-1}.$$

Similarly, $\int_0^1 \Delta(\nu)\,d\nu = \sum_{i=1}^{\infty} i\mu(i) = \frac{\tau}{I}\sum_{i=1}^{\infty} i\left(\frac{1}{1+\frac{\tau}{I}}\right)^i = \frac{1+\frac{\tau}{I}}{\frac{\tau}{I}}$, so that $M^{Dis} = \frac{1}{\Lambda\left(\frac{\tau}{I}\right)}\lambda^{-\int_0^1\Delta(\nu)d\nu} = \lambda^{-\frac{1+\frac{\tau}{I}}{\frac{\tau}{I}}}\frac{\lambda-1+\lambda\frac{\tau}{I}}{\frac{\tau}{I}}$.

equation are

$$\dot{p}_{\Delta}(a) = \begin{cases} (1 - p_0(a))\tau & \text{for } \Delta = 0 \\ -p_1(a)(I+\tau) & \text{for } \Delta = 1 \\ p_{\Delta-1}(a) I - p_{\Delta}(a)(I+\tau) & \text{for } \Delta \geq 2 \end{cases}.$$

The solution to this set of differential equations is given by

$$\begin{aligned} p_0(a) &= 1 - e^{-\tau \times a} \\ p_{i+1}(a) &= \left(\frac{1}{i!}\right) I^i a^i \left(e^{-(I+\tau)a}\right) \text{ for } i \geq 0. \end{aligned}$$

The distribution of markups conditional on survival is then

$$p^S_{i+1}(a) \equiv \frac{p_{i+1}(a)}{1 - p_0(a)} = \left(\frac{1}{i!}\right) I^i a^i \left(e^{-Ia}\right).$$

This is a Poisson distribution with parameter Ia, so that $E[\Delta|a] = Ia$. (14) then follows because $\ln(\mu) = \ln(\lambda)\Delta$.

The expected log markup by age: equation (15) From (2) we know that firm-level markups are given by$\mu_f = \frac{py_f}{wl_f} = \frac{1}{\frac{1}{n}\sum_{j=1}^{n_i} \lambda^{-\Delta_j}}$. Hence,

$$\ln(\mu_f) = -\ln\left(\frac{1}{n}\sum_{j=1}^{n_i} \lambda^{-\Delta_j}\right) \approx \ln(\lambda) \times \left[\frac{1}{N_f} \times \sum_{j=1}^{N_f} \Delta_j\right].$$

so that $E[ln(\mu_f)|\text{Age} = \text{a}] = \ln(\lambda) \times E_n\left[E\left[\frac{1}{n}\sum_{j=1}^{n} \Delta_j|\text{Age} = a, N = n\right]|\text{Age} = a\right]$. Define the random variable $B = \{0, 1, 2, ..., n\}$ by

$$B = \begin{cases} 0 & \text{if none of the } n \text{ products was the initial product of the firm} \\ k & \text{if product } k \text{ was the initial product of the firm} \end{cases}.$$

Then

$$E\left[\frac{1}{n}\sum_{j=1}^{n} \Delta_j|\text{Age} = a, N = n\right] = \sum_{k=0}^{n} E\left[\frac{1}{n}\sum_{j=1}^{n} \Delta_j|\text{Age} = a, N = n, B = k\right] \times P(B = k|\text{Age} = a, N = n).$$

To simplify notation I will simply denote the conditioning as a, n and k respectively. Then

$$E\left[\frac{1}{n}\sum_{j=1}^{n} \Delta_j|a, n\right] = E[\Delta_j|a, n, \text{ni}] + \frac{1 - P(B = 0|a, n)}{n}\left(E[\Delta_j|a, n, \text{i}] - E[\Delta_j|a, n, \text{ni}]\right). \tag{A-20}$$

where $E[\Delta|a, n, \text{ni}]$ denotes the conditional expected value of Δ conditional on the fact that the product is *not* an initial product and $\sum_{k=1}^{n} P(B = k|a, n) = 1 - P(B = 0|a, n)$. Now let us solve for $E[\Delta|a, n, \text{ni}]$, $E[\Delta|a, n, \text{i}]$ and $P(B = 0|a, n)$ in turn.

Recovering $E[\Delta|a,n,\mathbf{ni}]$ and $E[\Delta|a,n,\mathbf{i}]$ Let U denote the age of the product so that

$$E[\Delta|a,n,\text{ni}] = E_u\{E[\Delta|U=u]|a,n,\text{ni}\} = E_u\left\{\sum_{i=1}^{\infty} i \times p(i,u)|a,n,\text{ni}\right\}, \tag{A-21}$$

where the second line uses the fact that conditional on *product* age, no other characteristic matters and $p(i,u)$ is the probability of having a quality gap i conditional on the product having an age of u. As shown above this distribution follows a Poisson distribution of the form

$$p(i,u) = \left(\frac{1}{(i-1)!}\right)(Iu)^{i-1} \times \exp(-Iu).$$

Hence, $\sum_{i=1}^{\infty} i \times p(i,u) = Iu+1$. (A-21) therefore implies that

$$E[\Delta|a,n,\text{ni}] = 1 + I \times \int_{u=0}^{a} u \times f_{U|A,\text{ni}}(u|a,\text{ni})\,du, \tag{A-22}$$

where $f_{U|A,\text{ni}}$ is the density of the conditional age distribution of a product. In Section OA-1.2.1 in the Online Appendix I show that this density is given by

$$f_{U|A,\text{ni}}(u|a,\text{ni}) = \frac{\tau e^{-\tau\times u} + xe^{-(x+\tau)a}e^{xu}}{1-e^{-(x+\tau)a}}. \tag{A-23}$$

From (A-22) and (A-23) one can show that

$$E[\Delta|a,n,\text{ni}] = 1 + I \times \frac{\frac{1}{\tau}(1-e^{-\tau a}) - \frac{1}{x}e^{-\tau a}(1-\exp(-xa))}{1-\exp(-(x+\tau)a)}. \tag{A-24}$$

Turning to $E[\Delta|a,n,\text{i}]$, it is clear that the *initial* product of a firm of age a is simply a years old. Hence,

$$E[\Delta|a,n,\text{i}] = 1 + Ia. \tag{A-25}$$

Solving for $P(B=0|a,n)$ Note first that $P(B=0|a,n) = \frac{P(B=0,a,n)}{P(a,n)}$. We are going to construct $P(B=0,a,n)$. Let us denote this probability by $Q(n,t)$ where t is the age of the firm. This probability evolves according to the differential equation

$$\dot{Q}(n,t) = x(n-1)Q(n-1,t) + \tau(n+1)Q(n+1,t) - n(x+\tau)Q(n,t) + \tau(p(n+1,t) - Q(n+1,t)), \tag{A-26}$$

where $p(n,t)$ denotes the probability of having n products at time t. Note also that $\dot{Q}(0,t) = \tau p(1,t)$. Define the function

$$H_Q(z,t) \equiv \sum_{n=0}^{\infty} Q(n,t)z^n. \tag{A-27}$$

Then $\frac{\partial H_Q(z,t)}{\partial z} = \sum_{n=1}^{\infty} nQ(n,t) z^{n-1}$ and $\frac{\partial H_Q(z,t)}{\partial t} = \dot{Q}(0,t) + \sum_{n=1}^{\infty} \dot{Q}(n,t) z^n$. Using (A-26) it follows that,

$$\begin{aligned}
\frac{\partial H_Q(z,t)}{\partial t} &= \tau p(1,t) + \sum_{n=1}^{\infty} \left[x(n-1) \times Q(n-1,t) + \tau(n+1) \times Q(n+1,t) - n(x+\tau) Q(n,t)\right] z^n \\
&\quad + \tau \sum_{n=1}^{\infty} p(n+1,t) z^n - \tau \sum_{n=1}^{\infty} Q(n+1,t) z^n \\
&= \frac{\tau}{z}\left(H_P(z,t) - H_Q(z,t)\right) + \left(xz^2 - (x+\tau) z + \tau\right) \frac{\partial H_Q(z,t)}{\partial z},
\end{aligned}$$

where, as in (A-27), we defined

$$H_P(z,t) = \sum_{n=0}^{\infty} p(n,t) z^n, \tag{A-28}$$

Now define

$$\Psi(z,t) \equiv H_P(z,t) - H_Q(z,t). \tag{A-29}$$

Then

$$\dot{\Psi}(z,t) = \dot{H}_P(z,t) - \dot{H}_Q(z,t) = \left(xz^2 - (x+\tau) z + \tau\right) \frac{\partial \Psi(z,t)}{\partial z} - \frac{\tau}{z} \Psi(z,t), \tag{A-30}$$

where $\dot{H}_P(z,t) = \left(xz^2 - (x+\tau) z + \tau\right) \frac{\partial H_P(z,t)}{\partial z}$ follows the derivations in Klette and Kortum (2004). To solve for $\Psi(z,t)$ we need an initial condition. As every firm enters with a single product, we know that $p(1,t) = 1$ and $p(n,t) = 0$ for $n \neq 1$. Similarly, $Q(n,0) = 0$ for all n. Hence, (A-27), (A-28) and (A-29) imply that

$$\Psi(z,0) = \sum_{n=0}^{\infty} p(n,0) z^n - \sum_{n=0}^{\infty} Q(n,0) z^n = z, \tag{A-31}$$

which is the required initial condition. The solution to (A-30) with the initial condition in (A-31) is given by (see Section OA-1.2.2 in the Online Appendix for the proof)

$$\Psi(z,t) = \frac{(\tau - x) z \times e^{-\tau t}}{x(z-1) \times e^{-(\tau - x)t} - (xz - \tau)}. \tag{A-32}$$

From Klette and Kortum (2004, p. 1014) we know that $H_P(z,t)$ takes a similar form

$$H_P(z,t) = \frac{\tau(z-1) \times e^{-(\tau - x)t} - (xz - \tau)}{x(z-1) \times e^{-(\tau - x)t} - (xz - \tau)}.$$

(A-29) and (A-27) therefore imply that

$$H_Q(z,t) = \Psi(z,t) - H_P(z,t) = \frac{\tau(z-1) \times e^{-(\tau - x)t} - (xz - \tau) - (\tau - x) z \times e^{-\tau t}}{x(z-1) \times e^{-(\tau - x)t} - (xz - \tau)}. \tag{A-33}$$

From the definition of H_Q in (A-27) we can recover $Q(n,t)$ as the coefficients of the Taylor approximation around $z = 0$. In Section OA-1.2.3 in the Online-Appendix, I show that

$$Q(n,t) = \left(1 - \frac{\tau e^{-xt} - x e^{-\tau t}}{\tau - x}\right) \times p(n,t), \tag{A-34}$$

where $p(n,t)$ is described by $p(0,t) = \frac{\tau}{x}\gamma(t)$, $p(1,t) = (1-\gamma(t))(1-p(0,t))$ and $p(n,t) = \gamma(t)^{n-1} p(1,t)$ and the function $\gamma(t)$ is given by

$$\gamma(t) = \frac{x\left(1-e^{-(\tau-x)t}\right)}{\tau - x \times e^{-(\tau-x)t}}.$$

Equation (A-34) has the important implication that the *conditional* probability of not having an initial product at time t is independent of n, i.e.

$$P(\text{not initial}|t,n) = \frac{Q(n,t)}{p(n,t)} = 1 - \frac{\tau e^{-xt} - xe^{-\tau t}}{\tau - x}.$$

Hence,

$$1 - P(B=0|a,n) = \frac{\tau e^{-xa} - xe^{-\tau a}}{\tau - x}. \tag{A-35}$$

Note that $P(\text{not initial}|0,n) = 0$ and $\lim_{t\to\infty} P(\text{not initial}|t,n) = 1$ as required.

Substituting (A-24), (A-25) and (A-35) into (A-20) yields

$$E[a_P|a_f] \equiv E_n\left[E\left[\frac{1}{n}\sum_{j=1}^{n}\Delta_j|a,n\right]|a\right] = E[\Delta|a,\text{ni}] + (1-P(B=0|a)) \times (E[\Delta|a,\text{i}] - E[\Delta|a,\text{ni}]) \times \sum_{n=1}^{\infty}\frac{1}{n} f_{N|A}(n|a),$$

where $f_{N|A}(n|a)$ is the conditional distribution of n conditional on a. This object is given by $f_{N|A}(n|a) = \frac{p(n,a)}{1-p(0,a)} = \gamma(a)^{n-1} \times (1-\gamma(a))$. Hence,

$$E[ln(\mu)|a] = \ln\lambda \times (1 + I \times E[a_P|a_f]),$$

where

$$\begin{aligned} E[a_P|a_f] &= \frac{\frac{1}{\tau}(1-e^{-\tau a}) - \frac{1}{x}e^{-\tau a}(1-e^{-xa})}{1-e^{-(x+\tau)a}} + \left(a - \frac{\frac{1}{\tau}(1-e^{-\tau a}) - \frac{1}{x}e^{-\tau a}(1-e^{-xa})}{1-e^{-(x+\tau)a}}\right) \\ &\times \left(\frac{\tau e^{-xa} - xe^{-\tau a}}{x\left(1-e^{-(\tau-x)a}\right)}\right) \times ln\left(\frac{\tau - x \times e^{-(\tau-x)a}}{\tau - x}\right). \end{aligned}$$

This is the required expression in (15).

A-1.4 Proofs for Section 2.7

Consider the distribution of firms across the number of products they produce. Let $\omega(n) \equiv F \times \tilde{\omega}(n)$, where $\tilde{\omega}(n)$ denotes the measure of firms producing $n$products, i.e. $\sum_{n=1}^{\infty}\tilde{\omega}(n) = 1$. We know from Klette and Kortum (2004) that

$$\tilde{\omega}(n) = \frac{\frac{1}{n}\left(\frac{x}{\tau}\right)^{n-1}}{\sum_{j=1}^{\infty}\frac{1}{j}\left(\frac{x}{\tau}\right)^{j-1}}.$$

In a stationary equilibrium, the mass of entering and exiting firms has to be equal so that

$$F = \frac{z}{\tau} \times \sum_{j=1}^{\infty}\frac{1}{j}\left(\frac{x}{\tau}\right)^{j-1} = \frac{z}{x} \times \sum_{j=1}^{\infty}\frac{1}{j}\left(\frac{x}{\tau}\right)^{j} = \frac{z}{x} \times \ln\left(\frac{z+x}{z}\right) = \frac{1-\vartheta_x}{\vartheta_x}\ln\left(\frac{1}{1-\vartheta_x}\right),$$

where $\vartheta_x = \frac{x}{\tau}$. The entry rate is therefore given by

$$\text{Entry rate} = \frac{z}{F} = \frac{z}{\frac{z}{x} \times \ln\left(\frac{z+x}{z}\right)} = \frac{x}{\ln\left(\frac{z+x}{z}\right)}. \tag{A-36}$$

The share of products produced by firms with at most k products is given by

$$S_k = \sum_{n=1}^{k} F\tilde{\omega}(n)\, n = \left(\frac{z}{\tau} \times \sum_{j=1}^{\infty} \frac{1}{j}\left(\frac{x}{\tau}\right)^{j-1}\right) \times \frac{\sum_{n=1}^{k} \frac{1}{n}\left(\frac{x}{\tau}\right)^{n-1} n}{\sum_{j=1}^{\infty} \frac{1}{j}\left(\frac{x}{\tau}\right)^{n-1}} = \frac{z}{\tau} \times \frac{\tau}{x} \sum_{n=1}^{k} \left(\frac{x}{\tau}\right)^{n} = 1 - \vartheta_x^k.$$

To derive the the employment life cycle, i.e. equation (17), consider first the distribution of sales conditional on age. Note that $E[\ln l|a] = E\left[\ln\left(\frac{nY}{w\mu_f}\right)|a\right] = \ln\left(\frac{Y}{w}\right) + E[\ln n|a] - E[\ln \mu_f|a]$. To calculate $E[\ln n|a]$ note that the distribution of n conditional on age is given by $f_{N|A}(n|a) = \gamma(a)^{n-1} \times (1-\gamma(a))$. Hence,

$$E[\ln n|a] = \left(\frac{1-\gamma(a)}{\gamma(a)}\right) \sum_{n=1}^{\infty} \ln n \times \gamma(a)^n,$$

where $\gamma(t) = \frac{x\left(1-e^{-(\tau-x)t}\right)}{\tau - x \times e^{-(\tau-x)t}}$. It can also be shown that $\frac{\partial E[\ln n|a]}{\partial \gamma} > 0$, that $\frac{\partial \gamma(a)}{\partial \tau} < 0$ and that $\frac{\partial \gamma(a)}{\partial x} > 0$. Hence, $\frac{\partial E[\ln n|a]}{\partial x} > 0$ and $\frac{\partial E[\ln n|a]}{\partial \tau} < 0$.

A-1.5 Proof of Proposition 4

I prove the different parts in turn.

1. Write the equilibrium conditions in (A-13) and (A-14) as

$$1 = \frac{\tau}{\lambda\tau + I(\lambda-1)} \times \left(\frac{\lambda}{\lambda-1}\frac{\zeta}{\varphi_I} I^{\zeta-1}(\rho+\tau) + \zeta\frac{1}{\varphi_I} I^{\zeta}\right) + \frac{1}{\varphi_z} \times \tau - h(\varphi) = H^{\tau}(I,\tau) - h(\varphi) \quad \text{(A-37)}$$

$$\frac{1}{\varphi_z} = \frac{\zeta}{\varphi_I} I^{\zeta-1} + \frac{\frac{\zeta-1}{\varphi_I} I^{\zeta}}{\rho+\tau} + \frac{h(\varphi)}{\rho+\tau} = G^{\tau}(I,\tau). \quad \text{(A-38)}$$

It is easy to see that $\frac{\partial H^{\tau}}{\partial \tau} > 0$, $\frac{\partial G^{\tau}}{\partial \tau} < 0$ and $\frac{\partial G^{\tau}}{\partial I} > 0$. Also note that

$$\frac{\partial H^{\tau}}{\partial I} = \frac{\tau(\lambda-1)\frac{\zeta}{\varphi_I} I^{\zeta}\left[(\rho+\tau)\frac{\lambda}{(\lambda-1)I}\left(\frac{(\zeta-1)\lambda\tau}{(\lambda-1)I} + \zeta - 2\right) + \frac{\zeta\lambda\tau}{(\lambda-1)I} + \zeta - 1\right]}{[\lambda\tau + I(\lambda-1)]^2}.$$

The nominator is positive for $\zeta \geq 2$ and negative for $\zeta = 1$. Hence, that is some $\zeta \geq \overline{\zeta}$ such that $\frac{\partial H^{\tau}}{\partial I} > 0$. A sufficient condition is $\zeta \geq 2$. Hence, (A-37) implies a schedule $I^H(\tau)$, which decreasing and (A-38) implies a schedule $I^G(\tau)$, which is increasing. Furthermore, note that (A-15) implies that

$$\frac{\partial h(\varphi)}{\partial \varphi_x} = \frac{\zeta}{\zeta-1} h(\varphi) > 0$$
$$\frac{\partial h(\varphi)}{\partial \varphi_z} = -\frac{\zeta}{\zeta-1} h(\varphi) < 0.$$

Hence, an increase in φ_x or φ_z shifts the $I^H(\tau)$ curve up and the $I^G(\tau)$ curve down. This shows that higher entry costs and higher market barriers reduce τ.

2. Now consider the effect of φ_z and φ_x on ϑ_I. Consider again two equations (A-16) and (A-17). There I showed that these equations define the increasing schedule I^{FE} (from (A-17)) and the decreasing schedule I^{LM} (from (A-16)). The same argument then shows that entry costs and market barriers reduce ϑ_I.

3. From (9) it is immediate that higher entry costs increase x. Hence, higher entry costs increase ϑ_x.

4. Use (A-15) to write $\tau = \frac{\tau}{x} \times x = \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi)$. Then we can write (A-37) and (A-38) as

$$
\begin{aligned}
1 &= \frac{\frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi) \times \left(\frac{\lambda}{\lambda-1}\frac{\zeta}{\varphi_I} I^{\zeta-1}\left(\rho + \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi)\right) + \zeta\frac{1}{\varphi_I} I^{\zeta}\right)}{\lambda\frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi) + I(\lambda-1)} + \left(\frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1} - 1\right) h(\varphi) \\
\frac{1}{\varphi_z} &= \frac{\zeta}{\varphi_I} I^{\zeta-1} + \frac{\frac{\zeta-1}{\varphi_I} I^{\zeta}}{\rho + \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi)} + \frac{h(\varphi)}{\rho + \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi)}.
\end{aligned}
$$

Note that market barriers only enter through the h function. Again let the first schedule be $H\left(I, \vartheta_x^{-1}\right)$ and the second schedule be $G\left(I, \vartheta_x^{-1}\right)$. It is easy to see that $\frac{\partial G}{\partial I} > 0$ and $\frac{\partial G}{\partial \vartheta_x^{-1}} < 0$. Hence, $G\left(I, \vartheta_x^{-1}\right)$ implies un upward-sloping schedule $I^G\left(\vartheta_x^{-1}\right)$. It is also clear that $\frac{\partial H}{\partial \vartheta_x^{-1}} > 0$ as ϑ^{-1} is proportional to τ. Furthermore, under the same conditions as above, i.e. $\zeta \geq \overline{\zeta}$, we have $\frac{\partial H}{\partial I} > 0$. Hence, the H schedule defines a downward sloping locus $I^H\left(\vartheta_x^{-1}\right)$ in the $\left(I, \vartheta_x^{-1}\right)$ space. Now note that $\frac{\partial H}{\partial h} > 0$. As $\frac{\partial h}{\partial \varphi_x}$, an increase in φ_x will shift the $I^H\left(\vartheta_x^{-1}\right)$ schedule down. Similarly,

$$
\frac{\partial G}{\partial h} = \frac{\rho + \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi) - \left(\frac{\zeta-1}{\varphi_I} I^{\zeta} + h(\varphi)\right)\frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z}{\left(\rho + \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi)\right)^2} = \frac{\rho - \frac{\zeta-1}{\varphi_I} I^{\zeta}\frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z}{\left(\rho + \frac{1}{\vartheta_x}\frac{\zeta}{\zeta-1}\varphi_z h(\varphi)\right)^2},
$$

For ρ small, this implies that $\frac{\partial G}{\partial h} < 0$. Hence, an increase in φ_x will shift the $I^G\left(\vartheta_x^{-1}\right)$ up. This implies that an increase in ϕ_x will increase ϑ_x. Conversely, market barriers will reduce ϑ_x

5. The calibrated model in Section 3.3 is an example where the growth consequences are ambiguous.

A-1.6 The model with stochastic step size (Proof of Proposition 5)

In this section I derive the main results for the stochastic step size model. The detailed derivations are contained in Section OA-1.4 in the Online Appendix. Suppose that conditional on an innovation the step size of the quality increase is stochastic. Let the probability of climbing k rungs of the ladder be p_k with $\sum_{k=1}^{\infty} p_k = 1$.

The Equilibrium Value Function and the Equilibrium Conditions

As I show in Section OA-1.4 in the Online Appendix, the value function is still given by

$$
V_t\left(n, [\Delta_i]_{i=1}^n\right) = V_t^P(n) + \sum_{i=1}^{n} V_t^M(\Delta_i),
$$

where

$$
V_t^M(\Delta) = \frac{\pi_t(\Delta) - \pi_t(1) + (\zeta-1)\, c_I(I, \Delta)\, w_t}{\rho + \tau}
$$

and $V_t^P(n) = v_t n$, where

$$
v_t = \frac{\left(1 - \frac{1}{\lambda}\right) Y_t + (\zeta - 1)\frac{1}{\varphi_x \phi_e^{\zeta}} x^{\zeta} w_t}{\rho + \tau}.
$$

The optimal innovation and expansion rates are given by

$$\begin{aligned} I &= \left(\frac{E\left[1-\lambda^{-j}\right]\varphi_I \frac{Y_t}{w_t} - (\zeta-1)\frac{1}{\varphi_I} I^{\zeta}}{\zeta \quad \rho+\tau} \right)^{1/(\zeta-1)} \\ x &= \left(\frac{\varphi_x}{\varphi_e}\frac{1}{\zeta} \right)^{1/(\zeta-1)}. \end{aligned}$$

The free entry condition is given by

$$\frac{1}{\varphi_e} = \sum_{j=1}^{\infty} \left(\frac{V_t^P(1) + V_t^M(j)}{w_t} \right) p_j = \frac{1}{\rho+\tau} \left(\frac{Y_t}{w_t} + (\zeta-1)\frac{1}{\varphi_x} x^{\zeta} + \left((\zeta-1)\frac{1}{\varphi_I} I^{\zeta} - \frac{Y_t}{w_t} \right) \sum_j \lambda^{-j} p_j \right).$$

Together with the labor market condition, these equations fully determine the equilibrium.

The Distribution of Markups and First-Order Stochastic Dominance

The distribution of quality gaps $\nu(\Delta)$ solves the set of differential equations

$$\dot{\nu}(\Delta,t) = \begin{cases} -(\tau+I)\,\nu(\Delta,t) + I\sum_{j=1}^{\Delta-1}\nu(\Delta-j,t)\,p_j + \tau p_{\Delta} & \text{if } \Delta \geq 2 \\ \tau\left(p_1 - \nu(1,t)\right) - \nu(1,t)\,I & \text{if } \Delta = 1 \end{cases}.$$

The stationary distribution is therefore given by

$$\nu(j) = \frac{1}{1+\vartheta_I}\left(\sum_{m=1}^{j-1} \nu(m)\,p_{j-m} \right) + \frac{\vartheta_I}{1+\vartheta_I} p_j, \tag{A-39}$$

Define the cdf of quality gaps and hence markups as $\Phi(k) = \sum_{j=1}^{k}\nu(j)$. I now show that

$$\vartheta_H > \vartheta_L \rightarrow \Phi(k;\vartheta_H) > \Phi(k;\vartheta_L) \text{ for all } k.$$

To see this, define $\alpha = \frac{\vartheta}{1+\vartheta}$. α is increasing in ϑ. Write (OA-24) as

$$\nu(j) = (1-\alpha)\left(\sum_{m=1}^{j-1} \nu(m)\,p_{j-m} \right) + \alpha p_j.$$

The cdf Φ can be written as

$$\Phi(k) = \sum_{j=1}^{k}\nu(j) = (1-\alpha)\sum_{j=1}^{k}\sum_{m=1}^{j-1}\nu(m)\,p_{j-m} + \alpha\sum_{j=1}^{k} p_j = (1-\alpha)\sum_{m=1}^{k-1} p_{k-m}\Phi(m) + \alpha\sum_{j=1}^{k} p_j.$$

Let $\Phi(k;\alpha)$ denote the cdf as a function of α. Then

$$\begin{aligned} \Phi(k;\alpha_H) - \Phi(k;\alpha_L) &= (1-\alpha_H)\sum_{m=1}^{k-1} p_m \Phi(k-m;\alpha_H) + \alpha_H \sum_{j=1}^{k} p_j - (1-\alpha_L)\sum_{m=1}^{k-1} p_m \Phi(k-m;\alpha_L) - \alpha_L \sum_{j=1}^{k} p_j \\ &= (1-\alpha_H)\sum_{m=1}^{k-1} p_{k-m}\left[\Phi(m;\alpha_H) - \Phi(m;\alpha_L)\right] + (\alpha_H - \alpha_L)\left[p_k + \sum_{j=1}^{k-1} p_j \left(1 - \Phi(k-j;\alpha_L)\right) \right]. \end{aligned} \tag{A-40}$$

Now note that

$$\Phi(1;\alpha_H) - \Phi(1;\alpha_L) = (\alpha_H - \alpha_L)\, p_1 > 0.$$

Furthermore, (A-40) implies that

$$\Phi(m;\alpha_H) - \Phi(m;\alpha_L) > 0 \text{ for all } m < j \rightarrow \Phi(j;\alpha_H) - \Phi(j;\alpha_L) > 0$$

as $1 - \Phi(k-j;\alpha_L) > 0$ by Φ being a cdf. This shows that $\Phi(m;\alpha_H) - \Phi(m;\alpha_L) > 0$ for all m.

The Case of $p_n = \frac{1-\kappa}{\kappa} \times \kappa^n$

Suppose the step size is drawn from

$$p_n = \frac{1-\kappa}{\kappa} \times \kappa^n. \tag{A-41}$$

This distribution is parametrized by a single parameter κ. For $\kappa \rightarrow 0$, we recover the baseline model.

We now first show that the distribution of markups is again a pareto distribution. Using (OA-24) and (A-41), the density ν_j solves the equation

$$\nu_j = (1-\alpha)\left(\sum_{m=1}^{j-1} \nu_m \frac{1-\kappa}{\kappa}\kappa^{j-m}\right) + \alpha\frac{1-\kappa}{\kappa}\kappa^j,$$

where $\alpha = \frac{\vartheta}{1+\vartheta}$. Conjecture that $\nu_j = A^{j-1}\nu_1$ for $j \geq 2$. Substituting above yields

$$\begin{aligned}
\nu_j &= (1-\alpha)\left(\nu_1\frac{1-\kappa}{\kappa}\kappa^{j-1} + \sum_{m=2}^{j-1} A^{m-1}\nu_1\frac{1-\kappa}{\kappa}\kappa^{j-m}\right) + \alpha\frac{1-\kappa}{\kappa}\kappa^j \\
A^{j-1}\nu_1 &= \left[(1-\alpha)\frac{1-\kappa}{\kappa}\kappa^{j-1}\left(1 + \sum_{m=1}^{j-2}\left(\frac{A}{\kappa}\right)^m\right) + \kappa^{j-1}\right]\nu_1.
\end{aligned}$$

It is easy to show that $A = 1-(1-\kappa)\,\alpha = \frac{1+\kappa\vartheta}{1+\vartheta}$ solves this equation. Note that $\nu_1 = \frac{\vartheta}{1+\vartheta}p_1 = \frac{(1-\kappa)\vartheta}{1+\vartheta}$. Hence,

$$\nu_j = \left(\frac{1+\kappa\vartheta}{1+\vartheta}\right)^j \frac{\vartheta(1-\kappa)}{1+\kappa\vartheta},$$

The corresponding cdf is given by

$$\Phi(k) = \sum_{m=1}^{k}\nu_m = \nu_1\sum_{m=1}^{k}A^{m-1} = \nu_1\sum_{m=0}^{k-1}A^m = \nu_1\frac{1-A^k}{1-A} = 1 - \left(\frac{1+\kappa\vartheta}{1+\vartheta}\right)^k.$$

Hence,

$$P[\Delta \leq d] = 1 - e^{k\times\ln\left(\frac{1+\kappa\vartheta}{1+\vartheta}\right)}.$$

The distribution of markups is given by

$$\begin{aligned}
P[\mu \leq m] &= P\left[\lambda^{\Delta} \leq m\right] = P\left[\Delta \leq \frac{\ln m}{\ln\lambda}\right] = 1 - e^{\frac{\ln m}{\ln\lambda}\times\ln\left(\frac{1+\kappa\vartheta}{1+\vartheta}\right)} \\
&= 1 - m^{-\frac{1}{\ln\lambda}\ln\left(\frac{1+\vartheta}{1+\kappa\vartheta}\right)}.
\end{aligned}$$

Hence, the distribution is again pareto with shape parameter $\theta(\kappa) = \frac{1}{\ln\lambda}\ln\left(\frac{1+\vartheta}{1+\kappa\vartheta}\right)$. Because all aggregate wedges

are expressed in terms the pareto tail, all other results apply directly. To derive the expression for the aggregate growth rate $g = \frac{1}{1-\kappa}(I+\tau)\ln\lambda$, note that $g = (I+\tau)\ln\lambda\left(\sum_{n=1}^{\infty} kp_n\right)$. Then

$$\sum_{n=1}^{\infty} np_n = \sum_{n=1}^{\infty} n\frac{1-\kappa}{\kappa} \times \kappa^n = \frac{1}{1-\kappa}.$$

A-1.7 The model with CES preferences (Proof of Proposition 6)

In this section I prove the main results for the model with CES preferences. For detailed derivations I refer to Section OA-1.5 in the Online Appendix.

The static allocations can be derived by standard argument. The dynamic environment for own-innovation, entry and incumbent creative destruction is the same as in the baseline model. The only difference with respect to the baseline is that I assume that a fraction $(1-\delta)$ of creative destruction activities result in a "reset" of the quality of the destroyed product to the level λQ_t. This change is necessary to make the productivity distribution stationary. I discuss this in much more detail in Section OA-1.5 in the Online Appendix. All the aggregate implications independent of the parameter δ. The need for a stationary distribution of quality only arises when taking the model to the data. For continuity with the baseline model I still assume that the quality gap Δ after such a reset is still equal to unity.

The payoff-relevant state variable for a firm, which is producing n products is given by $[\Delta_i, q_i]_{i=1}^n$. The value function $V_t\left([\Delta_i, q_i]_{i=1}^n\right)$ therefore solves the HJB equation

$$\begin{aligned} & r_t V_t\left([\Delta_i, q_i]_{i=1}^n\right) - \dot{V}_t\left([\Delta_i, q_i]_{i=1}^n\right) \\ = & \sum_{i=1}^n \pi_t\left([\Delta_i, q_i]_{i=1}^n\right) + \sum_{i=1}^n \tau_t\left[V_t\left([\Delta_j, q_j]_{j\neq i}\right) - V\left([\Delta_i, q_i]_{i=1}^n\right)\right] \qquad \text{(A-42)} \\ & + \max_{[I_i]_{i=1}^n}\left\{\sum_{i=1}^n I_i\left[V_t\left([\Delta_j, q_j]_{j\neq i}[\Delta_i+1, \lambda q_i]\right) - V_t\left([\Delta_i, q_i]_{i=1}^n\right)\right] - \sum_{i=1}^n c^I\left(I_i, q_i\right) w_t\right\} \\ & + \max_X\left\{X\left[\delta\int_q V_t\left([\Delta_i, q_i]_{i=1}^n, 1, \lambda q\right) dF_t\left(q\right) + (1-\delta) V_t\left([\Delta_i, q_i]_{i=1}^n, 1, \lambda Q_t\right) - V_t\left([\Delta_i, q_i]_{i=1}^n\right)\right] - c^X\left(X, n\right) w_t\right\} \qquad \text{(A-43)} \end{aligned}$$

As before, I continue to assume each worker employed in entry activities generates a flow of φ_z of marketable ideas. For symmetry, a fraction δ of such ideas improve the existing quality of a randomly selected product by a step-size λ and a fraction $1-\delta$ "reset" the productivity to λQ_t. The free entry condition is therefore given by

$$\frac{1}{\varphi_z} w_t = \delta\int_q V_t\left(1, \lambda q\right) dF_t\left(q\right) + (1-\delta) V_t\left(1, \lambda Q_t\right). \qquad \text{(A-44)}$$

Suppose that $c^X\left(X, n\right)$ is as in the baseline model and that $c^I\left(I, q\right)$ is given by

$$c_t^I\left(I; \Delta, q\right) \quad = \quad \frac{1}{\varphi_I}\left(\frac{q}{Q_t}\right)^{\sigma-1} I^{\zeta}. \qquad \text{(A-45)}$$

In Section OA-1.5 in the Online Appendix I show that the value function $V_t\left([\Delta_i, q_i]_{i=1}^n\right)$ is given by

$$V_t\left([\Delta_i, q_i]_{i=1}^n\right) = \sum_{i=1}^n \frac{\psi\left(\Delta_i\right)}{\rho+\tau+(\sigma-1)g}\left(\frac{q_i}{Q_t}\right)^{\sigma-1}\frac{Y_t}{E[\mu^{1-\sigma}]} + \frac{1}{\rho+\tau}\frac{\zeta-1}{\varphi_x}\left(\frac{\varphi_x}{\varphi_z}\frac{1}{\zeta}\right)^{\frac{\zeta}{\zeta-1}} w_t n,$$

where $\psi(\Delta_i)$ is implicitly defined and depends only on Δ (and general equilibrium variables). I also show that the optimal innovation rate is given by

$$I(\Delta) = \left[(\psi(\Delta) - \alpha(\Delta)) \frac{1}{(\zeta - 1)} \frac{1}{\frac{1}{\varphi_I}} \left(\frac{L^P}{E[\mu^{-\sigma}]} \right) \right]^{\frac{1}{\zeta-1}},$$

where $\alpha(\Delta) = \left(1 - \frac{1}{\min\left\{\frac{\sigma}{\sigma-1}, \lambda^{\Delta}\right\}}\right) \min\left\{\frac{\sigma}{\sigma-1}, \lambda^{\Delta}\right\}^{1-\sigma}$. Hence, I is independent of q and constant along the BGP. Let $\nu_t(\Delta)$ be the mass of products with quality gap Δ. This distribution satisfies the differential equation

$$\frac{d\nu_t(\Delta)}{dt} = (I(\Delta - 1))\nu_t(\Delta - 1) - (\tau + I(\Delta))\nu_t(\Delta) \text{ for } \Delta \geq 2.$$

The law of motion for the mass of products with a quality gap of one is given by

$$\frac{d\nu_t(1)}{dt} = \tau - (I(1) + \tau)\nu_t(1).$$

Along a BGP this is distribution is stationary and given by

$$\nu(\Delta) = \frac{\tau}{I(\Delta)} \left(\prod_{j=1}^{\Delta} \frac{I(j)}{\tau + I(j)} \right). \tag{A-46}$$

Note that if $I(j) = I$ we have $\nu(\Delta) = \frac{\tau}{I}\left(\frac{I}{\tau+I}\right)^{\Delta} = \vartheta_I \left(\frac{1}{\vartheta_I+1}\right)^{\Delta}$ as in the baseline model. Given that markups are a one-to-one function of quality gaps (see (OA-25)), the distribution of markups is also stationary and only a function of τ and $\{I(\Delta)\}_{\Delta=1}^{\infty}$.

A-2 Appendix: Empirical Results

A-2.1 Measuring markups

To measure markups I closely follow the approach of De Loecker and Warzynski (2012). The crucial empirical object is the firms' labor share $s_{l,ft} = \frac{w_t l_{ft}}{p_f y_{ft}}$. As pointed out by De Loecker and Warzynski (2012), the level of production y_{ft} might contain both unanticipated shocks to and measurement error. Hence, they propose to consider a regression of

$$\ln y_{ft} = \phi(l_{ft}, k_{ft}, m_{ft}, z_{ft}) + \varepsilon_{ft}, \tag{A-47}$$

where $\phi(.)$ is estimated flexibly. Given the estimate $\hat{\phi}(.)$ one can recover an estimate of the measurement error $\hat{\varepsilon}_{ft}$ and form $s_{l,ft} = \frac{w_t l_{ft}}{p_f \frac{y_{ft}}{exp(\hat{\varepsilon}_{ft})}}$ (see De Loecker and Warzynski (2012, Equation 16)). Note that this correction is in terms of physical output. As in their application, I only have access to revenue and not physical output and hence I treat deflated sales as a measure of physical quantity. I therefore measure the cost share $s_{l,ft}$ as

$$s_{l,ft} = \frac{wl_{ft}}{va_{ft}/exp(\hat{\varepsilon}_{ft})},$$

where va_{ft} is observed value added, $\hat{\varepsilon}_{ft}$ is the residual from (A-47), where I take va_{ft} is the dependent variable and take $\phi(.)$ a second order polynomial in all (log) inputs and their interaction terms. For the specification with

	log employment					
Age	0.0930***	0.0814***		0.0923***	0.0811***	0.0317***
	(0.00193)	(0.00178)		(0.00248)	(0.00192)	(0.00210)
Entry			-0.187***	0.0466***		
			(0.00821)	(0.0101)		
Exit			-0.359***	-0.337***		
			(0.0106)	(0.0105)		
Industry FE	N	Y	Y	Y	Y	N
Firm FE	N	N	N	N	N	Y
N	76106	76106	64958	64958	59602	76106
R^2	0.035	0.218	0.212	0.231	0.239	0.927

Notes: Robust standard errors in parentheses with ***, **, and * respectively denoting significance at the 1%, 5% and 10% levels. I focus on the unbalanced panel of firms, who enter the market after 1990. I use the data from 1991 to 2000. All specifications include year fixed effects. 'Entry" and "Exit" are indicator variables for whether the firm enters (exit) the market in a given year. In column 5 I focus on the balanced panel, i.e. only consider firms that survive to the end of my sample period. The specifications with industry fixed effects control for industry affiliation at the 5 digit level. Column 5 contains firm fixed effects.

Table A-1: The employment life cycle in Indonesia

intermediate inputs instead of labor, the procedure is analogous.

A-2.2 The employment life cycle in Indonesia

The calibration uses the employment life cycle in Indonesia as an explicit calibration target. Focusing on the unbalanced panel of firms entering the economy after 1991, firms grow by about 8% a year. For the calibration I therefore use the estimated profile depicted in Figure 4 and target the log difference in employment for 7 year old firms. In Table A-1 I report additional regression results of predicting firm employment from firm age. The specification is exactly the same as (22), except that I do not control for firms' capital intensity. Column 3 shows that entrants and exiting firms are substantially smaller than the average firm. Column 4 shows that entrants are not too small given their age (in fact, they are slightly bigger) but that exiting firms are much smaller. This is exactly what the model predicts, because exiting firms are selected on n conditional on age, while entrants are not. Column 5 shows that the estimated age profile is quite similar once I condition on survival until the end of the sample. Finally, the last column controls for firm fixed effects. This lowers the age coefficient substantially.

A-2.3 The markup-size relationship

To quantify the relationship between firm-level markups and size, it is useful to derive an analytical expression analogous to Proposition 3. Note first that

$$E\left[ln\left(\mu_f\right)|N=n\right] = log\left(\lambda\right) \times E_a\left[E\left[\frac{1}{n}\sum_{j=1}^{n}\Delta_j|\text{Age}=a, N=n\right]|N=n\right].$$

Above I already showed that

$$E\left[\frac{1}{n}\sum_{j=1}^{n}\Delta_j|\text{Age}=a, N=n\right] = E\left[\Delta_j|a,n,\text{ni}\right] + \frac{1-P\left(B=0|a,n\right)}{n}\left(E\left[\Delta_j|a,n,\text{i}\right] - E\left[\Delta_j|a,n,\text{ni}\right]\right),$$

where $E[\Delta_j|a,n,\text{ni}]$ is given in A-24, $E[\Delta_j|a,n,\text{i}]$ is given in A-25 and $1-P(B=0|a,n)$ is given in A-35. In particular, none of these objects depend on n. Hence,

$$\begin{aligned} E[ln(\mu_f)|N=n] &= \int_a \left[E[\Delta|a,\text{ni}] + \frac{(1-P(B=0|a))}{n}(E[\Delta|a,\text{i}]-E[\Delta|a,\text{ni}])\right] f(a|n)\,da \\ &= 1 + I \times \int_a \left[g(a,x,\tau) + \frac{(1-P(B=0|a))}{n}(a-g(a,x,\tau))\right] f(a|n)\,da \end{aligned} \tag{A-48}$$

where

$$g(a,x,\tau) = \frac{\frac{1}{\tau}(1-e^{-\tau a}) - \frac{1}{x}e^{-\tau a}(1-\exp(-xa))}{1-\exp(-(x+\tau)a)},$$

and $f(a|n)$ is the distribution of age conditional on size, which is given by

$$f(a|n) = \frac{(1-\gamma(a))\gamma(a)^{n-1}\left(1-\frac{\tau}{x}\gamma(a)\right)}{\frac{1}{x}\frac{1}{n}\left(\frac{x}{\tau}\right)^n}, \tag{A-49}$$

where $\gamma(a) = \gamma(a) = \frac{x\left(1-e^{-(\tau-x)a}\right)}{\tau-xe^{-(\tau-x)a}}$.[36]

The expressions above fully determine the average log markup as a function of n. In Figure 8 I show the results for the calibrated model. The left panel shows the average markup as a function of size, i.e. (A-48). In the right panel I show the stationary firm size distribution.

Figure 8 shows that the average markup is increasing in size - at least for the vast majority of firms. The very top part of the sales distribution where markups are declining in size is of course closely related to the top of the age distribution where markups are declining in age - see Figure 1. Figure 8 also shows that the quantitative effect firm size on the average markup is very small. Increasing size by one log point (say from 1 to 2) increases the average markup by 0.6%. This is the elasticity reported in the main text. The firm-level data shows a stronger relationship between markups and size. In Table A-2 I report the results of regression of log markups (columns 1 and 2) and log TFPR (columns 3 and 4) on log sales. The estimated elasticity is consistently estimated to be around 0.2, i.e. much larger than in the model.

[36] To derive (A-49), note first that the mass of firms with n products is given by

$$M_n = \frac{\theta}{n}\left(\frac{1}{1+\theta}\right)^n = \frac{\frac{z}{x}}{n}\left(\frac{1}{1+\frac{z}{x}}\right)^n = \frac{z}{x}\frac{1}{n}\left(\frac{x}{\tau}\right)^n.$$

The probability of having n products at time t when born at time $t-a$ is given by $p_n(a)$, where

$$p_n(a) = (1-\gamma(a))\gamma(a)^{n-1}\left(1-\frac{\tau}{x}\gamma(a)\right).$$

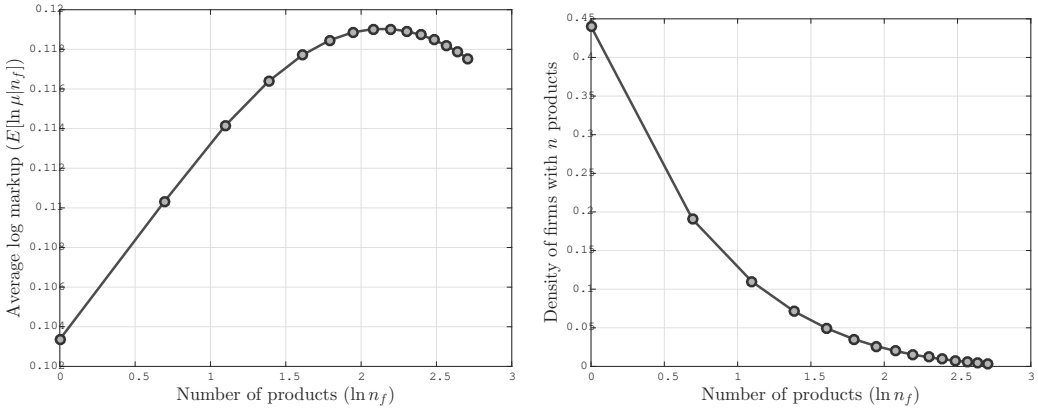
Each period z firms enter. Hence,

$$M_n = \int_{a=0}^{\infty} zp_n(a)\,da.$$

Then conditional distribution is therefore given by

$$f(a|n) = \frac{zp_n(a)}{M_n} = \frac{(1-\gamma(a))\gamma(a)^{n-1}\left(1-\frac{\tau}{x}\gamma(a)\right)}{\frac{1}{x}\frac{1}{n}\left(\frac{x}{\tau}\right)^n},$$

which is the expression in (A-49).



Notes: In the left panel I depict the average (log) markup as a function of $\ln n$. In the right panel I depict the stationary distribution of sales. The endogenous flow rates (I, x, z) and the parameter λ correspond to the calibration of the baseline model.

Figure 8: Markups by Size

	Dependent variable			
	log Markups ($\ln \mu_f$)		log TFPR ($\ln py/(k^\alpha l^{1-\alpha})$)	
log sales	0.192***	0.168***	0.163***	0.229***
	(0.000925)	(0.00109)	(0.00119)	(0.00112)
$\ln k/l$		0.0547***		-0.242***
		(0.00134)		(0.00137)
N	176958	138953	122578	122578
R^2	0.306	0.311	0.153	0.354

Notes: Robust standard errors in parentheses with ***, **, and * respectively denoting significance at the 1%, 5% and 10% levels. All specifications include year fixed effects and 5-digit industry fixed effects. $ln\,(k/l)$ denotes the (log) capital-labor ratio at the firm level.

Table A-2: Markups, TFPR and Size