ECONOMETRICA

HETEROGENEOUS MARKUPS, GROWTH,
AND ENDOGENOUS MISALLOCATION

MICHAEL PETERS
Department of Economics, Yale University

HETEROGENEOUS MARKUPS, GROWTH, AND ENDOGENOUS MISALLOCATION

MICHAEL PETERS
Department of Economics, Yale University

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 are stochastically replaced by more efficient competitors. Creative destruction therefore has pro-competitive effects because faster churn gives firms less time to accumulate market power. In an application to firm-level data from Indonesia, the model predicts that, relative to the United States, misallocation is more severe and firms are substantially smaller. To explain these patterns, the model suggests an important role for frictions that prevent existing firms from entering new markets. Differences in entry costs for new firms are less important.

KEYWORDS: Creative destruction, markups, firm dynamics.

1. INTRODUCTION

FIRMS' MARKET POWER IS an important determinant of static allocative efficiency. Similarly, the welfare consequences of particular policies directly depend on whether and how firms' markups change in 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 in a model of imperfect product markets where firms compete à la Bertrand, engage in non-competitive pricing, and charge variable markups. By endogenizing the extent of market power in a canonical model of growth, this paper offers a unified perspective on the relationship between markups, misallocation, the process of firm dynamics, and aggregate growth.

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 because 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 for a given product are therefore increasing as long as the product is produced by the same firm. Once the firm is replaced by a new producer, markups are "reset" as Bertrand competition intensifies. These two forces of markup-increasing quality improvements by firms in their existing

products and markup-reducing product churning through creative destruction shape the equilibrium distribution of markups.

The model can be solved analytically and allows for a precise theoretical characterization of 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 that I call the *churning intensity*. This statistic is simply the rate of creative destruction relative to the rate at which firms increase their market power. Intense churning compresses the cross-sectional distribution of markups because new firms replace existing producers quickly and keep monopoly power limited. If, by contrast, the extent of churning is limited, the distribution of markups has a fat tail because incumbent firms have ample time to accumulate market power.

Second, I show that this endogenous Pareto tail fully determines the macroeconomic implications of market power. More specifically, both aggregate TFP and the labor share—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: a higher churning intensity reduces misallocation and increases aggregate TFP and the labor share.

Despite its parsimony, the model has rich testable implications for the empirical patterns of markups, age, and size across firms. At the product level, the model predicts that markups follow a distinct life-cycle pattern: conditional on survival, markups increase stochastically as a result of the firm's productivity-enhancing investments. This "own-innovation channel" raises the average markup of old firms relative to young firms. At the firm level, two counteracting effects are at play because firms expand into new products and lose existing products to other firms. This "creative destruction channel" lowers the average markup of old firms, because markups for new products are, on average, lower than those on the old products the firm loses. 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.

Motivated by the recent literature on misallocation in developing countries, I apply the theory to firm-level panel data from Indonesia. 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 (sometimes also referred to as "TFPR"), and then quantifies its effect on cross-country differences in aggregate TFP. Because markups are a form of firm-specific wedges that are determined endogenously, the model provides a way to quantify the share of TFPR dispersion, which is due to markups. When calibrated to firm-level panel data from Indonesia, the model implies 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%.

To better understand the underlying determinants of misallocation and firm size, I consider a specific counterfactual exercise. I start from the premise that a developing economy such as Indonesia might suffer from frictions that hamper firms' ability to enter new product markets. Such frictions could be related to policies such as size requirements and lengthy approval processes, or they could be technological in nature, whereby the costs of breaking into new markets are higher in developing countries. The model highlights that such frictions come in two flavors: expansion costs make breaking into new product markets costly for incumbents; 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 U.S. grow faster than their Indonesian counterparts, but the rates of entry are roughly similar. While these two moments imply that both entry and expansion costs are lower in the U.S., I find that the difference in expansion costs for existing firms is twice as important. The reason is simple: compared to firms in Indonesia, firms in the U.S. grow faster conditional on survival, thus making the average firm larger. Lower entry costs have 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 U.S. This has implications for the distribution of firm size, misallocation, and aggregate growth. Relative to Indonesia, the aggregate importance of small firms declines by 75% and average firm size more than doubles. Moreover, the increase in churning reduces misallocation by roughly one-third; that is, it increases allocative efficiency by 0.3%. The implications for the growth rate are subtle. Higher entry and expansion costs reduce creative destruction, which lowers the equilibrium growth rate. At the same time, by increasing the survival probabilities for existing firms, such barriers increase the incentives for firms to invest in productivity growth in their existing products. This increases the equilibrium growth rate, albeit at the cost of higher markups. In my calibration, these two effects essentially cancel each other. 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.   My theory builds on models of endogenous growth in the Schumpeterian tradition of Aghion and Howitt (1992), Grossman and Helpman (1991), and in particular Klette and Kortum (2004). This framework is analytically attractive and can rationalize many salient features of the data (Lentz and Mortensen (2008), Akcigit and Kerr (2018)). It has been used to study industrial policies (Atkeson and Burstein (2019), Acemoglu, Akcigit, Alp, Bloom, and Kerr (2018)), to quantify the importance of managerial delegation (Akcigit, Alp, and Peters (2020)), to analyze the optimal protection of intellectual property rights (Acemoglu and Akcigit (2012)), and to measure the sources of U.S. growth (Garcia-Macia, Hsieh, and Klenow (2019)). I show how to extend this framework in a tractable way to endogenize the distribution of markups. This extra margin not only generates 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 economic environment (see, e.g., Foster, Haltiwanger, and Syverson (2008), De Loecker and Warzynski (2012), or De Loecker, Goldberg, Khandelwal, and Pavcnik (2016)). Recently, interest in the macroeconomic implications of market power has been growing. De Loecker and Eeckhout (2017) and Autor, Dorn, Katz, Patterson, and Van Reenen (2017), for example, argued that markups have been increasing and that highly profitable "superstar" firms have become more important for the aggregate economy. A growing literature also analyzes the consequences of the declining dynamism in the U.S. on markups, the aggregate labor share, and misallocation (Decker, Haltiwanger, Jarmin, and Miranda (2014), Haltiwanger, Decker, and Jarmin (2015), Akcigit and Ates (2019), Aghion, Bergeaud, Boppart, Klenow, and Li (2019), Peters and Walsh (2019), Edmond, Midrigan, and Xu (2018)). The theory proposed in this paper is qualitatively consistent with these trends as it predicts 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, which treats misallocation as an exogenous firm-specific wedge (see, e.g., Hsieh and Klenow (2009), Restuccia and Rogerson (2008), Bartelsman, Haltiwanger, and Scarpetta (2013), David and Venkateswaran (2019), and the survey article by Hopenhayn (2014)). A recent literature has also studied how such exogenous distortions affect firm growth and entry (see, e.g., Hsieh and Klenow (2014), Bento and Restuccia (2017), Da-Rocha, Tavares, and Restuccia (2017), Fattal Jaef (2018) and Buera and Fattal Jaef (2016)). In this paper, I take the opposite approach. Misallocation is fully endogenous and hence emerges as an equilibrium outcome together with the size distribution of firms.[1]

Finally, a growing literature in international trade stresses the importance of markups. On the theory side, Bernard, Eaton, Jensen, and Kortum (2003), Melitz and Ottaviano (2008), and Atkeson and Burstein (2008) developed models that generate heterogeneous markups. The effects of markup heterogeneity and misallocation on the welfare gains from trade were explicitly discussed in Edmond, Midrigan, and Xu (2015), Epifani and Gancia (2011), Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2019), De Blas and Russ (2015), and Holmes, Hsu, and Lee (2014). In contrast to my model, all of 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 Supplemental Material (Peters (2020)) contains most proofs of the theoretical results and additional details regarding the empirical analysis.

2. THEORY

A measure one of infinitely lived households supply their unit-time endowment inelastically. Individuals have preferences over the unique consumption good given by

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

This numeraire final good is a Cobb–Douglas composite of a continuum of differentiated products

$$\ln Y_t = \int_{i=0}^{1} \ln\left(\sum_{f \epsilon S_{it}} y_{fit}\right) di,$$

where y_{fit} is the quantity of product i bought from firm f, and S_{it} denotes the set 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 analytically convenient. In Section 2.7, I show how the analysis can be extended to a more general setting.

---

[1]Although numerous theories of input-based misallocation have been proposed (e.g., imperfect capital markets (Buera, Kaboski, and Shin (2011), Moll (2014), Midrigan and Xu (2014)), information frictions (David, Hopenhayn, and Venkateswaran (2016)), or adjustment costs (Asker, Collard-Wexler, and Loecker (2014))), this paper is the first to structurally explore the role of monopoly power as a source of misallocation.

Firms can produce multiple products, and the only source of heterogeneity across firms is their product-specific efficiency. 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. Firms take aggregate prices as given, but compete à la Bertrand with other producers offering the same product. 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, and (iii) productivity increases by current producers in markets they already serve. Whereas 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 a competitive edge relative to other firms, gives rise to heterogeneous markups across producers, and provides the link between growth, misallocation, and the process of firm dynamics.

2.1. *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 drop the time subscripts if doing so does not cause confusion. Because production takes place with constant returns, firms compete in prices, and different brands of product i are perceived as perfect substitutes, only the most productive firm within a product market will be active in equilibrium. However, the presence of competing producers (even though they are less efficient) constrains the leading firm to engage in limit pricing. Letting q_i denote the productivity of the most efficient firm, 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 by the productivity advantage over competing firms.[2]

To derive employment and profits at the product level, the Cobb–Douglas assumption on consumers' preferences implies that sales are equalized, that is, $p_i y_i = Y$. Then

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

Note that both employment l_i and profits π_i depend only on the markup μ_i and not on the level of efficiency q_i, which makes the characterization of the model extremely tractable.

---

[2]The fact that markups are fully determined from limit pricing makes the analysis tractable. In Section A-1.1 in the Supplemental Material, I discuss why the model would be less tractable if firms were to instead compete à la Cournot.

To aggregate these allocations to the firm level, let N_f be the set of products that firm f produces and $n_f = |N_f|$ be the number. Total employment of firm f, l_f, is then given by

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

Here, μ_f is the markup at the firm level, which is simply the harmonic mean 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 n_f. Conversely, higher markups μ_f reduce firm employment, holding the number of products fixed.

The economy aggregates in a transparent way. Letting L_{Pt} denote the total mass of production workers, (2) implies

$$L_{Pt} = \int_f l_{ft}\, df = \frac{Y_t}{w_t} \int_f \sum_{i \in N_f} \mu_{it}^{-1}\, df = \frac{Y_t}{w_t} \times \left(\int \mu^{-1}\, dG_t(\mu) \right),$$

where $G_t(\mu)$ denotes the cross-sectional distribution of markups at time t. Similarly, given that the final good is the numeraire, equilibrium wages are

$$w_t = Q_t \times \exp\left(\int \ln \mu^{-1}\, dG_t(\mu) \right),$$

where $\ln Q_t = \int \ln q\, dH_t(q)$ is the usual CES efficiency index and $H_t(q)$ is the distribution of productivity q. Aggregate output is therefore given by

$$Y_t = Q_t \mathcal{M}_t L_{Pt}, \quad \text{where } \mathcal{M}_t \equiv \frac{\exp\left(\int \ln \mu^{-1}\, dG_t(\mu) \right)}{\int \mu^{-1}\, dG_t(\mu)}. \qquad (3)$$

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 misallocation. It is easy to show that $\mathcal{M} \leq 1$ and that $\mathcal{M} = 1$ if and only if markups are equalized. Hence, aggregate TFP depends on the *dispersion* of markups. Whereas a common proportional increase in markups leaves the degree of misallocation unchanged, higher markup dispersion reduces allocative efficiency and hence aggregate TFP.

Monopoly power affects not only aggregate TFP but also factor prices. In particular, equilibrium wages are distorted relative to their social marginal product and satisfy

$$\frac{w_t L_{Pt}}{Y_t} = \Lambda_t, \quad \text{where } \Lambda_t \equiv \int \mu^{-1}\, dG_t(\mu). \qquad (4)$$

In contrast to $\mathcal{M}$, the labor share Λ depends on the level of markups, because higher markups reduce wages and increase aggregate profits. Note also that the canonical case of constant markups as generated by a CES demand system with differentiated products is a special case: TFP is identical to its competitive counterpart (i.e., $\mathcal{M}_t = 1$), but factor prices are lower (i.e., $\Lambda_t = \mu^{-1} < 1$).

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 depend only on the marginal distribution of markups $G_t(\mu)$.[3] I now construct this marginal distribution as an endogenous outcome from firms' innovation decisions so that misallocation, factor shares, and growth are jointly determined in equilibrium.

2.2. *Dynamics: Innovation and Creative Destruction*

Both the aggregate physical productivity 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

$$\mu_{it} = \frac{q_{it}}{q_{it}^F} = \frac{\lambda^{r_{it}}}{\lambda^{r_{it}^F}} = \lambda^{r_{it} - r_{it}^F} \equiv \lambda^{\Delta_{it}}, \tag{5}$$

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 productivity gap Δ and the equilibrium markup μ.

The current productivity of product i, q_{it}, can increase in three ways: (i) a new firm can enter the market for product i with a superior technology ("creative destruction by entrants"), (ii) an existing firm that 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 in order to gain additional monopoly power ("own-innovation"). Although these three sources of growth all improve the current frontier productivity by a single step from q_{it} to λq_{it}, they have very different allocative consequences. In the case of own-innovation, the equilibrium markup *increases* by a factor λ as the productivity gap increases from Δ to $\Delta + 1$. By contrast, when productivity growth is due to creative destruction, the equilibrium markup *decreases* by a factor $\lambda^{\Delta(i,t)-1}$, as the quality gap declines from Δ to unity.[4]

To characterize the equilibrium rates of own-innovation, incumbent creative destruction, and entry, I need to solve for the value function, as all these choices are forward-looking. In principle, the state of a firm consists of the qualities and quality gaps of all its products, $\{q_{it}\}_{i=1}^n$ and $\{\Delta_{it}\}_{i=1}^n$. To economize on notation, I refer to these sets as $[q_i]$ and $[\Delta_i]$. Equations (1) and (5) imply that equilibrium profits in product i are given by $\pi_{it} = (1 - \lambda^{-\Delta_{it}})Y_t$; that is, they depend only on the productivity gap. I therefore restrict attention to equilibria where firm behavior depends only on the payoff-relevant state variables $(n, [\Delta_i])$.

I adopt the usual stochastic formulation whereby firms choose the flow rates of increasing the productivity of existing products and of expanding into a novel, randomly selected

---

[3]This is a consequence of the Cobb–Douglas structure. In Section 2.7, 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.

[4]In Section 2.7, 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 product experiences both entry and a productivity improvement by the current producer.

product. I denote the rate of own-innovation on existing products by $[I_i]$ and the rate of expansion by $[x_i]$. The associated cost function (denoted in units of labor) is given by $\Gamma([x_i, I_i]; n, [\Delta_i])$.[5] Optimal behavior is then described by the value function $V_t(n, [\Delta_i])$, which is given by

$$r_t V_t(n, [\Delta_i]) = \sum_{j=1}^{n} \underbrace{\pi_t(\Delta_j)}_{\text{Flow profits}} + \underbrace{\dot{V}_t(n, [\Delta_i])}_{\text{Capital gain}} - \sum_{j=1}^{n} \underbrace{\tau_t \big[V_t(n, [\Delta_i]) - V_t(n-1, [\Delta_i]_{i \neq j}) \big]}_{\text{Creative destruction}}$$

$$+ \max_{[x_j, I_j]} \left\{ \sum_{j=1}^{n} \underbrace{I_j \big[V_t(n, [[\Delta_i]_{i \neq j}, \Delta_j + 1]) - V_t(n, [\Delta_i]) \big]}_{\text{Own-innovation}} \right.$$

$$\left. + \sum_{j=1}^{n} \underbrace{x_j \big[V_t(n+1, [[\Delta_i], 1]) - V_t(n, [\Delta_i]) \big]}_{\text{Expansion}} - \underbrace{\Gamma([x_i, I_i]; n, [\Delta_i]) w_t}_{\text{Innovation costs}} \right\}.$$

The value of a firm consists of four parts. First, there is the total flow payoff, which is simply the sum of profits across all markets. Second, the firm could benefit from a capital gain in case the value function increases over time. Third, the firm could lose any of its n existing products to other firms, which happens at the endogenous rate of creative destruction τ. Finally, there is an option value term, that captures the possibility of increasing the markup in each of its products through own-innovation and of expanding into new products. Note that the productivity gap in novel products is always equal to unity; that is, new products have low markups.

While this expression for V_t looks daunting, it turns out that V_t admits a simple closed-form solution under the assumption of the following additively separable cost function $\Gamma(\cdot)$ that is consistent with balanced growth:

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

Here, $\zeta > 1$ 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 and institutional determinants such as bureaucratic requirements to produce a particular product.[6]

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 into a single market with a unitary quality gap, the free-entry condition is given by

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

---

[5]For notational simplicity, I formulate the firm's problem in terms of the expansion rates *per own product* x_i. An equivalent formulation assumes that the firm as a whole chooses an expansion rate $X = \sum x_i = nx$.

[6]The term $\lambda^{-\Delta}$ in $c^I(.)$ implies that innovations are easier the bigger the productivity advantage Δ. Intuitively: per-period profits are given by $(1 - \lambda^{-\Delta})Y$ and hence are 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 $\lambda^{-\Delta}$ is exactly the right normalization to balance those effects. Note that firms generate a high productivity gap only when they have multiple innovations *in a row*. Hence, this specification posits that firms can build on their own innovations of the past.

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 given by

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

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

2.3. *The Stationary Equilibrium*

Given this setup, 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 unit aggregate labor endowment

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

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 choose $[y_{it}, c_t]_{it}$ to maximize utility, (iii) firms choose $[I_{it}, x_{it}, p_{it}]$ 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.

Although firms' market power is endogenous, the stationary equilibrium can be characterized analytically. Two properties are important for this result. First, the theory admits closed-form solutions for the value function V_t and incumbents' innovation behavior. This is the content of Proposition 1. Second, the distribution of markups, which is required to calculate the misallocation wedge $\mathcal{M}$ and the labor wedge Λ, can also be characterized analytically and I do so in Section 2.4.

PROPOSITION 1: *Consider the setup described above. Suppose $\rho > \frac{\zeta-1}{\zeta}(\frac{1}{\zeta}\frac{\varphi_x}{\varphi_z})^{1/(\zeta-1)}$. Then there exists a unique stationary equilibrium, where*:
1. *The value function is given by*

$$V_t(n, [\Delta_i]) = \sum_{i=1}^{n} V_t(\Delta_i) = V_t^P n + \sum_{i=1}^{n} V_t^M(\Delta_i), \tag{6}$$

where

$$V_t^P = \frac{\pi_t(1) + (\zeta-1)\dfrac{x^{\zeta}}{\varphi_x} w_t}{\rho + \tau} \quad and \quad V_t^M(\Delta) = \frac{\pi_t(\Delta) - \pi_t(1) + (\zeta-1)\lambda^{-\Delta}\dfrac{I^{\zeta}}{\varphi_I} w_t}{\rho + \tau}.$$

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

$$x = \left(\frac{\varphi_x}{\varphi_z} \frac{1}{\zeta} \right)^{\frac{1}{\zeta - 1}}, \tag{7}$$

and the innovation rate I solves

$$I = \left(\frac{\lambda - 1}{\lambda} \frac{1}{\rho + \tau} \left(\frac{\varphi_I}{\zeta} \frac{Y_t}{w_t} - \frac{\zeta - 1}{\zeta} I^\zeta \right) \right)^{\frac{1}{\zeta - 1}}. \tag{8}$$

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*

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

PROOF: See Section A-1.2 in the Supplemental Material. The condition $\rho > \frac{\zeta - 1}{\zeta} \times (\frac{1}{\zeta} \frac{\varphi_x}{\varphi_z})^{1/(\zeta - 1)}$ is sufficient for the free-entry condition to be satisfied. *Q.E.D.*

Proposition 1 establishes that the economy admits a unique stationary equilibrium that can be characterized analytically. The value of the firm, $V_t(n, [\Delta_i])$, has an intuitive structure. First, it is additive across products. Second, the value of producing a given product with quality gap Δ 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, that is, the option value of being able to expand into 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 market power. It consists of the flow value of being able to charge higher markups $(\pi_t(\Delta) - \pi_t(1))$, 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 are optimal innovation, entry, and expansion decisions, which are constant across products and time. Equation (7) shows that the rate at which incumbents enter new markets has a simple closed-form expression and depends on the efficiency of incumbents' creative destruction (φ_x) relative to entrants' (φ_z). The optimal extent of own-innovation I depends on two endogenous aggregate variables: the rate of creative destruction τ and the 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 to earn 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 $\tau = z + x$.[7] The aggregate growth

---

[7]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 (6), (7), and (8) are six equations in the six unknowns $(x, \tau, I, \frac{V_t}{w_t}, \frac{Y_t}{w_t}, L_{Pt})$. In Section A-1.2, I show how these can be reduced to a system of two equations in two unknowns.

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 size λ, g is proportional to the sum of creative destruction τ and firms' rate of own-innovation I.

2.4. *The Cross-Sectional Distributions of Markups*

The distribution of markups is endogenous and determined in equilibrium. To construct this distribution, recall that markups depend only on the distribution of quality gaps Δ across products. The cross-sectional distribution of markups is therefore fully characterized by $\{\nu_t(\Delta)\}_{\Delta=1}^\infty$, where $\nu_t(\Delta)$ denotes the measure of products with quality gap Δ at time t. These measures solve the set of differential equations

$$\dot{\nu}_t(\Delta) = \begin{cases} -(\tau + I)\nu_t(\Delta) + I\nu_t(\Delta - 1) & \text{if } \Delta \geq 2, \\ \tau\big(1 - \nu_t(1)\big) - I\nu_t(1) & \text{if } \Delta = 1, \end{cases} \tag{9}$$

where $\dot{\nu}_t(\Delta)$ denotes the time derivative. Intuitively, product i can leave the state Δ in two ways: 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 state Δ 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 for which the producing firm is replaced enter this state.

Equation (9) is the key equation to characterize the equilibrium distribution of markups. Three properties are noteworthy. First, the distribution is fully determined from the two endogenous variables (I, τ) and is hence jointly determined with the economy-wide growth rate g. Second, the distribution of firm size is not required to solve for the distribution of markups across products, because all firms innovate and expand at constant rates I and x. Finally, creative destruction is pro-competitive: whereas productivity growth by existing producers I increases markups, market churning through creative destruction τ shifts the distribution of markups downward. This suggests that creative destruction reduces misallocation and firms' own-innovation efforts increase misallocation. 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*

$$\theta = \frac{\ln(1 + \vartheta)}{\ln \lambda}, \quad \text{where } \vartheta = \frac{\tau}{I}.$$

1. *The stationary distribution of markups is given by*

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

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

$$\mathcal{M} = e^{-1/\theta}\frac{1 + \theta}{\theta} \quad \text{and} \quad \Lambda = \frac{\theta}{1 + \theta}. \tag{10}$$

PROOF: See Section A-1.3 in the Supplemental Material. *Q.E.D.*

Proposition 2 contains the main theoretical result of this paper: the equilibrium 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. 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, by contrast, churning is of little importance, the resulting distribution of markups has a fat tail, and both the average markup and its dispersion are large. Because I and τ are determined endogenously and hence depend on parameters and policies, markups and misallocation will also endogenously respond. This is the crucial difference relative to Bernard et al. (2003), who generate a Pareto distribution of markups from firms' exogenous productivity draws.[8]

The macroeconomic consequences of market power are fully summarized by the two sufficient statistics $\mathcal{M}$ and Λ. These also depend only on ϑ and have the closed-form expressions given in (10).[9] It is easy to verify that both $\mathcal{M}$ and Λ are increasing in ϑ. 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.[10]

The mechanism that generates the endogenous Pareto tail in my model is akin to the city-size dynamics in Gabaix (1999). Markups *within a product* have an intuitive life-cycle interpretation: as long as the current producer is not 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-1.4 of the Supplemental Material, the distribution of the number of quality gaps Δ conditional on the length of time a product has been produced by a particular firm, which I refer to as "product age" a_P, is Poisson distributed with parameter $I a_P$. This implies that the average log markup of a product conditional on being produced by the same firm for a_P years is given by

$$E[\ln \mu | \text{product age} = a_P] = \ln \lambda (1 + I a_P), \tag{11}$$

that is, it is increasing in age at a rate proportional to I. Hence, conditional on not being replaced, the distribution of markups continuously shifts outward as incumbent firms engage in productivity improvements to ratchet up their monopoly power. The higher is I, the faster this process of accumulation.

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 on the degree of creative destruction τ. If τ is high, it is rare to see firms serving a particu-

---

[8]De Blas and Russ (2015) analyzed a version of Bernard et al. (2003), where the number of firms is finite and the distribution of markups responds to trade policy.

[9]Note that Δ is not a continuous variable, but only takes integer values. For simplicity, I treat markups as continuous when calculating $\mathcal{M}$ and Λ from $G(\mu)$. See Section A-1.3 in the Supplemental Material for the closed-form expressions for the discrete case.

[10]Note also that the standard deviation of log markups is given by θ^{-1} and is hence also decreasing in ϑ. Hence, as stressed in the literature on misallocation, the dispersion in log TFPR co-moves negatively with aggregate TFP.

lar product market for a long time. The interplay of these two processes leads to a Pareto distribution of markups in the long run.[11]

2.5. *Markup Dynamics at the Firm Level*

Equation (11) characterizes 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 products. This 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 because they are the only firms with the potential of having had enough time to accumulate market power 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 have also had ample time to expand into new markets and lose products from their portfolio. And as markups for new, "marginal" products are lower than markups for the average product that firms lose, this "creative destruction channel" tends to lower the extent to which markups increase in age.

To see this more clearly, suppose firms were to never horizontally expand (i.e., $x = 0$) and hence never serve more than a single market. In that case, the age of the firm a_f directly corresponds to the age of the product a_P and the average log markup by *firm* age is also given by (11). Allowing firms to expand horizontally into new markets breaks this tight link between markups and firm age. Nevertheless, one can still derive an analytical characterization of the markup life-cycle at the firm level.

PROPOSITION 3: *The average firm-level markup* $\ln \mu_f$ *as a function of age is given by*

$$E[\ln \mu_f | firm\ age = a_f] = \ln \lambda \times (1 + I \times E[a_P|a_f]), \tag{12}$$

where

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

and

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

PROOF: See Section A-1.4 in the Supplemental Material. $\qquad$ *Q.E.D.*

The life-cycle profile of markups in (12) has the same structure as (11), except that the mapping between firm age a_f and product age a_P is more complicated. In particular, it depends on both the rate of incumbent expansion x and creative destruction τ. The possibility of firms breaking into new markets implies $E[a_P|a_f] \leq a_f$. Moreover, it is easy to verify that $\lim_{x \to 0} E[a_P|a_f] = a_f$, so that (11) emerges as a special case.

---

[11]To see this intuitively, suppose (11) were to hold deterministically, that is, $\ln \mu = \ln \lambda + \ln \lambda \times Ia$. Then, $P[\mu > \mu_0] = P\left[a < \frac{1}{I} \frac{\ln(\mu_0/\lambda)}{\ln \lambda}\right] = (\frac{\mu_0}{\lambda})^{-\frac{1}{\ln \lambda} \frac{\tau}{I}}$, which is a Pareto distribution. Jones and Kim (2018) exploit a similar structure to argue that creative destruction limits income inequality.
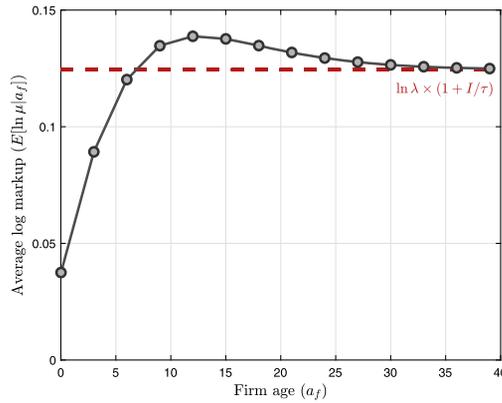
FIGURE 1.—THE LIFE-CYCLE OF MARKUPS. *Notes*: The figure displays the expected log markup by age, that is, $E[\ln \mu_f | a_f]$ given in (12).

The life-cycle profile characterized in (12) is depicted in Figure 1. Surprisingly, the relationship is *non-monotone*. The reason is the following. Recall that a given product is creatively destroyed at flow rate τ. Hence, the average survival time for a given product is $1/\tau$. For the set of very old firms, this is exactly the average age of a given product in their portfolio as $\lim_{a_f \to \infty} E[a_P | a_f] = 1/\tau$. The limiting average markup for old firms (which is displayed in the red dashed line in Figure 1) is therefore given by $\ln \lambda (1 + I/\tau)$. Note that, like for the cross-sectional distribution of markups characterized in Proposition 2, the churning intensity $\vartheta = \tau/I$ again emerges as a crucial determinant for the level of markups. If the churning intensity ϑ is high, the average markup of old firms is low.

The average markup for younger firms deviates from this level due to selection. For young firms, the age of the products they sell is obviously *negatively* selected: a two-year-old producer cannot possibly 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 that are expected to increase. Interestingly, once firms become sufficiently old, the expected age of the products they sell is *positively* selected. In particular, the firm may still own 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 hang on to its initial product goes to zero.

Proposition 3 also highlights how the life-cycle of markups is shaped by firms' innovation choices (I, x) and the rate of creative destruction τ. First, it is immediate that (for a given age) the average markup is increasing in I because $E[a_P | a_f]$ is independent of I. 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 and in the rate of creative destruction τ.[12] Although both τ and x *decrease* the extent of life-cycle markup growth, the economics are very different. A higher rate of creative destruction τ, holding x fixed, reduces markup growth through higher churning as firms have less time to accumulate market power in the products they own. By contrast, a higher expansion rate x, holding τ fixed, reduces the average markup through a composition effect: if firms enter novel markets frequently, only a small fraction of their sales is accounted for by old, high-markup products.

---

[12]For visual clarity, I focus on the early part of the life-cycle, where markups are monotone in age, which is the empirically relevant case for my application.
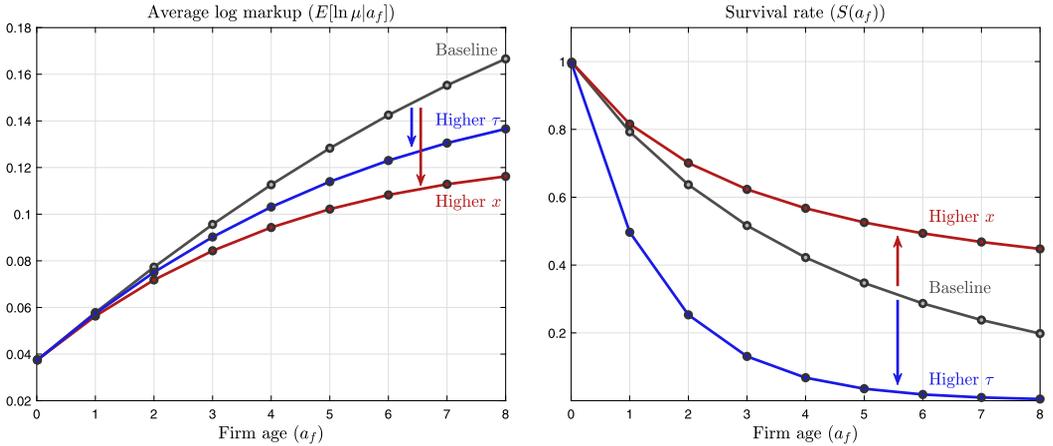
FIGURE 2.—THE LIFE-CYCLE OF MARKUPS AND FIRM SURVIVAL. *Notes*: The figure displays the expected log markup as a function of age (see (12)) in the left panel and the cumulative survival probability (see (13)) in the right panel. The baseline model is shown in gray. The red (blue) line refers to an increase in incumbent expansion x (creative destruction τ).

This suggests that a higher expansion rate x reduces markups and increases allocative efficiency. However, as shown in Proposition 2, 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, the share of firms surviving until age a_f, $S(a_f)$, is given by

$$S(a) = 1 - \frac{\tau}{x}\gamma(a), \tag{13}$$

where $\gamma(a)$ is given in Proposition 3. As seen in the right panel of Figure 2, a higher rate of incumbent creative destruction x increases the number of surviving firms at every age bin. Hence, while a faster expansion rate reduces the expected markup for a given age, it also increases the share of old firms, which have higher markups on average. In the cross-section, these two effects exactly cancel each other, rendering the distribution of markups independent of x conditional on τ. By 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 shifts the age distribution towards young, low-markup firms.

2.6. *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. Second, the stochastic process of firms losing and gaining products is the same as in Klette and Kortum (2004). The process of sales dynamics therefore takes exactly the same form: 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 of large firms.

Importantly, and in contrast to Klette and Kortum (2004), sales and employment are no longer proportional, because firm employment is also affected by the firm's average

markup.[13] Interestingly, the determinants of the markup and sales distribution can be neatly separated. Whereas the distribution of markups depends only on the churning intensity τ/I, the distribution of sales can be fully characterized by the endogenous statistic x/τ, that is, the share of creative destruction that is due to incumbent firms. In particular, as shown in Section A-1.5 in the Supplemental Material, the number of active firms F is given by $F = \frac{1-\frac{x}{\tau}}{\frac{x}{\tau}} \ln(\frac{1}{1-\frac{x}{\tau}})$ and the share of aggregate sales accounted for by firms with at most n products, Ω_n, is given by $\Omega_n = 1 - (\frac{x}{\tau})^n$. 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 horizontal expansion x and vertical own-innovation I. In particular, as I show in Section A-1.5 in the Supplemental Material, the life-cycle of firm size $E[\ln l_f | a_f]$ is given by

$$E[\ln l_f | a_f] - E[\ln l_f | 0] = \underbrace{(1 - \gamma(a)) \sum_{j=1}^{\infty} \ln j \times \gamma(a)^{j-1}}_{\text{Sales growth}} - \underbrace{\ln \lambda (I \times E[a_P | a_f])}_{\text{Markup growth}}, \qquad (14)$$

where $\gamma(a)$ and $E[a_P | a_f]$ 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, because markups are constant. By contrast, if firms could not expand horizontally, they would only produce a single product, sales would not be a function of age, and firm employment would decline along the life-cycle as markups increase. Hence, the speed at which firms increase their markups relative to their size identifies the importance of horizontal expansion relative to vertical productivity improvements in their existing products.

2.7. *Theoretical Extensions*

The baseline model laid out above can be solved analytically and cleanly highlights the core economic mechanism. This tractability of course required stringent assumptions. Of particular importance seem to be the restrictions to a constant step size and a unitary elasticity of demand. In this section, I discuss in more detail to what extent these assumptions are consequential. Detailed derivations are contained in Section A-1.6 in the Supplemental Material.

Stochastic Step Size. Consider first the assumption of a constant step size. It turns out that this assumption is easy to dispense with, and 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 $p_j = P[\tilde{k} = j]$ and $\sum_{j=1}^{\infty} p_j = 1$.

---

[13]Another difference is that my model predicts a deviation from the exact proportionality between R&D spending and sales. The R&D intensity, that is, 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). 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).

The baseline model is the special case with $p_1 = 1$. 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 4: *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*

$$\nu(\Delta) = \frac{1}{1+\vartheta}\left(\sum_{m=1}^{\Delta-1}\nu(m)\,p_{\Delta-m}\right) + \frac{\vartheta}{1+\vartheta}\,p_{\Delta}\quad \text{for } \Delta = 1, 2, 3, \ldots$$

Hence, the churning intensity $\vartheta = \tau/I$ is still a sufficient statistic for the distribution of quality gaps. Moreover, a higher churning intensity ϑ decreases the distribution of markups in a first-order stochastic dominance sense.

2. *Suppose $p_j = \frac{1-\kappa}{\kappa}\kappa^j$, where $\kappa < 1$. Define $\theta(\kappa) \equiv \frac{1}{\ln\lambda}\ln(\frac{1+\vartheta}{1+\kappa\vartheta})$. The distribution of markups and the aggregate misallocation measures $\mathcal{M}$ and Λ are given by*

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

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

PROOF: See Section A-1.6 in the Supplemental Material. *Q.E.D.*

Proposition 4 shows that the results from the baseline model extend in a straightforward way. The distribution of markups is still fully determined from the churning intensity ϑ and creative destruction is pro-competitive. The special case of $p_n \propto \kappa^n$ is particularly instructive as the distribution of markups is again Pareto with shape parameter $\theta(\kappa)$. As expected, holding ϑ constant, the shape parameter is decreasing in κ: a more dispersed exogenous step-size distribution results in a more dispersed markup distribution in equilibrium. As in the baseline model, an increase in ϑ raises the Pareto tail. Because the sufficient statistics $\mathcal{M}$ and Λ depend only on the Pareto tail, the formulas from Proposition 2 directly translate. The baseline model is nested as the case $\kappa = 0$.

While these results suggest that higher values of κ induce more misallocation in equilibrium, I show below that this is not the case quantitatively. Specifically, once the model is calibrated to match the same moments as the baseline economy, the value of κ is inconsequential. I discuss this in more detail in Section 3.4 below.

CES Preferences. A more fundamental assumption concerns the unitary demand elasticity. Suppose the final good was not a Cobb–Douglas aggregate, but instead took the more general CES form $Y_t = (\int_i y_{it}^{(\sigma-1)/\sigma}\,di)^{\sigma/(\sigma-1)}$, where y_{it} is the amount of variety i, which as before can be produced by multiple firms. While I relegate a detailed analysis to the Supplemental Material, 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, in the more general CES case, equilibrium markups are no longer necessarily determined by Bertrand competition, but firms with a sufficiently large productivity advantage might prefer to charge the usual CES markup instead of the limit price. Formally, the equilibrium markup is given by

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

Second, the misallocation term $\mathcal{M}_t$ and the labor wedge Λ_t now take the form

$$\mathcal{M}_t = \frac{\left(\int \mu(\Delta)^{1-\sigma}\left(\frac{q}{Q_t}\right)^{\sigma-1} dF_t(q, \Delta)\right)^{\frac{\sigma}{\sigma-1}}}{\int \mu(\Delta)^{-\sigma}\left(\frac{q}{Q_t}\right)^{\sigma-1} dF_t(q, \Delta)} \text{ and } \Lambda_t = \frac{\int \mu(\Delta)^{-\sigma}\left(\frac{q}{Q_t}\right)^{\sigma-1} dF_t(\Delta)}{\int \mu(\Delta)^{1-\sigma}\left(\frac{q}{Q_t}\right)^{\sigma-1} dF_t(\Delta)}, \quad (15)$$

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

$$\pi_t(q, \Delta) = \left(1 - \frac{1}{\mu(\Delta)}\right)\mu(\Delta)^{1-\sigma} q^{\sigma-1} w_t^{1-\sigma} Y_t.$$

These equations highlight why the case of $\sigma = 1$ is particularly tractable. First, the limit price is always binding so that $\mu = \lambda^{\Delta}$ for all Δ. Second, if $\sigma > 1$, the misallocation terms $\mathcal{M}_t$ and Λ_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. Third, both the level of quality q and the quality gap Δ determine profits π_i and hence are state variables for the firm's dynamic programming problem.

It turns out that one can still make theoretical progress in analyzing this more general case. Suppose that, as in Atkeson and Burstein (2010), the cost of own-innovation scales at the same rate in q as firm profits, that is,

$$c_t^I(I; \Delta, q) = \left(\frac{q}{Q_t}\right)^{\sigma-1}\frac{1}{\varphi_I}I^{\zeta}. \quad (16)$$

Under this assumption, the model is consistent with Gibrat's law; that is, growth rates are independent of firm size, at least for large firms.[14] In this context, one can derive the following characterization of a stationary equilibrium.

PROPOSITION 5: *Consider the model with CES demand and let the equilibrium be stationary. Then*:

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

---

[14]The cost function in (16) does not depend on the productivity gap Δ. This is for simplicity. As long as the cost function takes the form $c_t^I(I; \Delta, q) = (\frac{q}{Q_t})^{\sigma-1}c(I, \Delta)$, with $c(I, \Delta)$ being convex in I, one can show that the optimal rate of own-innovation is independent of q.

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

$$\nu(\Delta) = \vartheta(\Delta) \prod_{m=1}^{\Delta} \frac{1}{1+\vartheta(\Delta)}, \quad where \; \vartheta(\Delta) = \frac{\tau}{I(\Delta)}. \tag{17}$$

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

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

The main result of Proposition 5 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 (16), 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. The distribution of quality gaps is therefore determined by a set of differential equations akin to (9) in the baseline model, which has the solution (17). As before, only the churning intensities $\{\vartheta(j)\}_j$ are required to solve for the distribution of quality gaps and hence markups. If I were constant, (17) is exactly the same solution as in the baseline model. Because markups are still fully determined from firms' quality advantage Δ, the main economic insight from the baseline model is preserved in this more general environment.

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.[15] In particular, the misallocation terms $\mathcal{M}$ and Λ follow directly from (15) and (17). These expressions highlight how the demand elasticity σ affects the aggregate losses of misallocation $\mathcal{M}$. Holding $\mu(\Delta)$ and the distribution $\nu(\Delta)$ 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, a higher demand elasticity mechanically reduces markup dispersion because, holding the distribution of Δ constant, more products will charge a markup of $\frac{\sigma}{\sigma-1}$. Additionally, changes in σ also affect firms' innovation and entry incentives and hence the endogenous distribution $\nu(\Delta)$. In Section 3.4, I show quantitatively that the results 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

---

[15]The distribution of q is required to calibrate the model. 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. In fact, the quality distribution in the baseline model is not stationary, even though firm-level employment, sales, and profits are.

calibrated model to quantify the importance of heterogeneous 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). These data have also been used in Amiti and Konings (2007) 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, that is, the time period prior to the Indonesian financial crisis. My final sample has about 180,000 observations.[17]

The Statistik Industri data focus on large formal producers and therefore have a size threshold of 20 employees. In the context of a developing economy such as Indonesia, this sample of firms is heavily selected. Hsieh and Olken (2014), for example, analyzed data from the Indonesian economic census and found that the share of firms with fewer than 10 workers is essentially indistinguishable from 100%. At the same time, the firms in the Statistik Industri data are sufficiently large to account for roughly 40% of total employment. Table I contains some descriptive statistics and shows that the average plant has about 140 employees and that the firm size distribution is skewed: while the median plant in the survey has 45 employees, the 90th percentile of the distribution is 350. Compared to the U.S. manufacturing sector, plants in Indonesia are of course still small. In the U.S., 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.

This focus on large producers has advantages and disadvantages. On the positive side, these data cover firms for which considerations of productivity improvements and strategic pricing are more relevant. The majority of micro-firms in Indonesia are arguably subsistence entrepreneurs, who are unlikely to engage in such activities and who do not com-

TABLE I

THE MANUFACTURING SECTOR IN INDONESIA[a]

Firm Size Distribution				Entrants			Exiting Firms		
	Quantiles				Share of			Share of	
Mean	25%	50%	90%	Entry Rate	Empl.	Sales	Exit Rate	Empl.	Sales
143	27	45	351	10.4%	5.0%	3.9%	8.2%	4.4%	3.3%

[a]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 8 contain the entry and exit rate. Columns 6 and 9 (7 and 10) report the employment (sales) share of entering and exiting firms. All results are simple averages over the time period of the sample.

---

[16]To be absolutely precise, the data are collected at the plant level. As the majority of plants are reported to be single-branch entities, in what follows, I refer to each plant as a firm.

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

pete in the same product markets as large, formal employers.[18] Additionally, the data have a panel dimension, which allows me to use the information contained in the dynamics of markups and firm size to calibrate the structural parameters.

The main drawback of this size-based selection criterion is that it complicates the measurement of entry and exit, because I only observe firms appearing and disappearing from the data. Table I shows that, on average, 10.4% of firms enter and 8.2% of firms exit the data in a given year. Naturally, these firms are much smaller than the average firm; 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 the theory). To map these moments to the theory, note that if the relevant product markets are the ones 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 definition 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 explicitly.

3.2. *The Markup Life-Cycle of Indonesian Firms*

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

Measuring Markups. To measure markups, I follow the approach pioneered by Jan De Loecker (De Loecker et al. (2016), De Loecker and Warzynski (2012), De Loecker (2011)) and hence relegate most of the details to the Supplemental Material. 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 exploit only the growth of markups over the life-cycle as a calibration moment, I do not need to estimate the level of markups. Hence, 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 firm's cost-minimization problem imply that the markup satisfies the equation

$$\mu_f = \alpha_{l,f} \times s_{l,f}^{-1}, \tag{18}$$

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. In my model, the output elasticity of labor is unity, so that (18) implies (2). Note that the derivation of (18) did not use any information on the structure of demand or how firms compete.[19]

---

[18]Mounting evidence suggests that "stagnant" entrepreneurs are important in developing economics; see, for example, Schoar (2010), Hurst and Pugsley (2012), and Akcigit, Alp, and Peters (2020). That these firms do not compete in the same product markets is argued in La Porta and Shleifer (2009) and La Porta and Shleifer (2014).

[19]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

If α_l is known, one could directly infer firms' markups from their observed labor shares. If α_l is not known, but assumed to be constant, (18) 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

$$\ln s_{l,ft}^{-1} = \delta_s + \delta_t + u_{ft}, \tag{19}$$

that is, $\ln \hat{\mu}_{ft} = \hat{u}_{ft}$. 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 (18) using the estimated $\hat{\alpha}_l$.[20] However, I also consider richer specifications, where I explicitly control for firms' input choices such as 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.

The Life-Cycle of Markups. In Figure 3, I show that—as predicted by the theory—markups increase in age, at least for the majority of firms. Specifically, I focus on all firms that entered the economy after 1990, estimate the relative markup $\hat{u}_{ft}$ from (19), and then calculate the average markup by cohort age, relative to entering firms. As in the theory, there is attrition as firms exit the market, and the size of the markets reflects the number of surviving firms in the cohort. This schedule therefore refers exactly to the expression characterized in Proposition 3 and displayed in Figure 1.
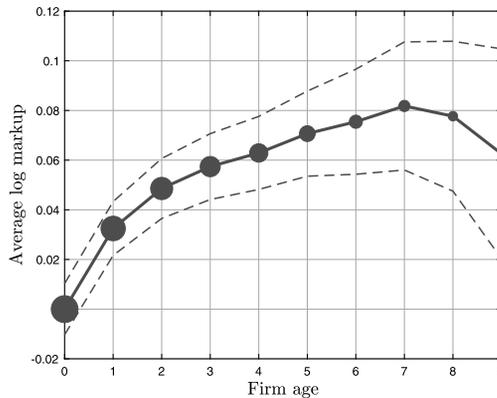

FIGURE 3.—THE LIFE-CYCLE OF MARKUPS IN INDONESIA. *Notes*: I focus on the unbalanced panel of firms entering the economy after 1990. I calculate log markups within 5-digit-industry-year cells, then calculate the average by the age of the cohort and normalize log markups of entering cohorts to zero. Because 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.

---

the labor share of the entire workforce at firm f, $s_{l,f}^{\text{Total}}$, is given by $s_{l,f}^{\text{Total}} = (1 + \frac{w}{Y}\frac{1}{\varphi_I}I^\zeta)\mu_f^{-1} + \frac{w}{Y}\frac{1}{\varphi_x}x^\zeta$. Because $\frac{w}{Y}\frac{1}{\varphi_x}I^\zeta$ and $\frac{w}{Y}x^\zeta$ are constant, the variation in $s_{l,f}^{\text{Total}}$ across firms and along the life-cycle is entirely driven by the variation in μ_f^{-1}.

[20]Furthermore, my data are standard in the sense that they do 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.

TABLE II

THE LIFE-CYCLE OF MARKUPS IN INDONESIA[a]

	Dependent Variable: Log Markup			
Age	0.0137	0.0103	0.0106	0.0108
	(0.00122)	(0.00117)	(0.00116)	(0.00127)
Industry FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Control for k/l		$\ln k/l$	non-param.	$\ln k/l$
Balanced panel				✓
N	55,212	55,212	55,212	42,434
R^2	0.219	0.287	0.305	0.298

[a] Robust standard errors in parentheses. I focus on the unbalanced panel of firms that enter the market after 1990. I use the data from 1991 to 2000. Columns 2 and 4 control for the log capital-labor ratio $\ln k/l$. Column 3 controls non-parametrically by including 50 fixed effects for 50 quantiles of $\ln k/l$. Column 4 focuses on the balanced panel. Industry fixed effects control for industry affiliation at the 5-digit level.

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

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

$$\ln(\mu_{ft}) = \delta_t + \delta_s + \beta \times \text{age}_{ft} + \varrho \times \ln(k_{ft}/l_{ft}) + h'_{ft}\gamma + \varepsilon_{ft}, \tag{20}$$

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 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), induced by systematic differences in production technologies within 5-digit industries. Doing so reduces the estimated age coefficient slightly. In column 2, I simply include the log of firms' capital-labor ratio, and in column 3, I control for capital intensity in a non-parametric way by including 50 fixed effects for 50 quantiles of the distribution of capital-labor ratios. In column 4, I directly control for selection by conditioning on survival. In the theory, no selection exists, because 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. However, markups are increasing over time even for those firms that do survive until the end of the sample.

In Section A-2.2 in the Supplemental Material, I present additional robustness checks for these results. In particular, I show that the results do not substantially depend on whether 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[21] and provide additional evidence for firms' factor shares to indeed reflect markups rather than other input distortions such as credit constraints.

3.3. *Calibration*

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. The model is parsimonious. Given a rate of time preference ρ, which I set exogenously, the theory is fully parameterized 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.

As shown in the expression for the life-cycles of firm size and markups, all micro-moments related to the process of firm dynamics depend only 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 or ρ. I then use the equilibrium conditions to find the required structural parameters yielding (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 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. (2018) and assume $\zeta = 2$ for my baseline results and provide robustness.

To identify the four structural parameters $(\varphi_I, \varphi_z, \varphi_x, \lambda)$, I use four moments, all of which have closed-form expressions in the theory. First, I target the average markup of 7-year-old firms relative to that of entrants, given in (12). As shown in Figure 3, markups increase by 0.08 log points over a 7-year horizon. Second, I match the observed life-cycle of employment, that is, the average employment of 7-year-old firms relative to that of entrants, given in (14). Empirically, firms in Indonesia increase their employment by roughly 0.5 log points in the first 7 years of their life. Third, I target the entry rate of 10.4% reported in Table I. In the model, the entry rate is given by

$$\text{Entry Rate} = \frac{z}{F} = \frac{x}{\ln\left(\frac{z+x}{z}\right)}.$$

Recall that F denotes the number of firms in equilibrium. Finally, I target a rate of aggregate productivity growth $g = \ln \lambda (\tau + I)$ of 3%. 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 boils down to solving four nonlinear equations. In Table III, 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 IV contains a sensitivity matrix and reports the change in equilibrium outcomes and moments for a 5% increase in the respective

---

[21]Markups distort all factors within the firm equally, so that (18) should also hold for the share of sales going to materials. The results are similar to the ones reported in Table II when firms' 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.

TABLE III

CALIBRATION: MOMENTS AND STRUCTURAL PARAMETERS[a]

Targets	Data	Model	Calibrated Parameters			Outcomes	
Markup life-cycle	0.082	0.082	φ_I	Innovation efficiency	8.896	I	0.561
Empl. life-cycle	0.494	0.494	φ_x	Expansion efficiency	1.077	x	0.207
Entry rate	10.4%	10.4%	φ_z	Entry efficiency	2.590	z	0.032
Agg. growth	3%	3%	λ	Step size	1.038	τ	0.240
–	–	–	ζ	Cost function	2		
–	–	–	ρ	Discount rate	0.05		

[a]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 7-year-old firms. 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.

structural parameters. The results conform well to the economic intuition of the theory. First, an increase in the efficiency of innovation (φ_I), expansion (φ_x), or entry (φ_z) raises own-innovation I, incumbent creative destruction x, and entry z, respectively, and both φ_x and φ_z negatively affect incumbent own-innovation, because the higher rate of creative destruction increases the effective discount rate of existing firms. That neither φ_I nor λ affects the equilibrium level of creative destruction by incumbents is apparent from (7). 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, reported in the lower panel of Table IV, 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. 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 portfolios at a faster rate. By 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

TABLE IV

SENSITIVITY MATRIX[a]

	Change in …				
	φ_I	φ_x	φ_z	λ	Initial Level
Effects on Endogenous Outcomes					
Rate of own-innovation (I)	4.0%	−1.1%	−0.9%	1.9%	0.561
Incumbent 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

[a]The table reports the effect of a 5% change in the relative innovation efficiencies (φ_I, φ_x, φ_z) and the quality increase ($\lambda - 1$) on the endogenous outcomes (top panel) and the equilibrium moments (lower panel).

intuitive: the rate of growth, the entry rate, and the markup life-cycle increase and the slope of the employment life-cycle declines.

Non-Targeted Moments. Figure 4 contains a set of outcomes that 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). Although the model is calibrated to match only the data for 7-year-old firms, it captures the general age pattern for markups and employment. In the lower panels, I focus on the dynamic patterns of exit (left panel) and the cross-sectional size distribution (right panel). Through the lens of the model, which is stationary, I can measure survival either from the cross-section at time t or by following a cohort through its life-cycle. In the lower left panel, I depict the model's implication for $S(a)$ (see equation (13)) and the corresponding data measured both from the cross-section in 2000 and from the evolution of the cohort that entered in 1991. Two properties stand out. First, the survival probabilities estimated from the cross-section and
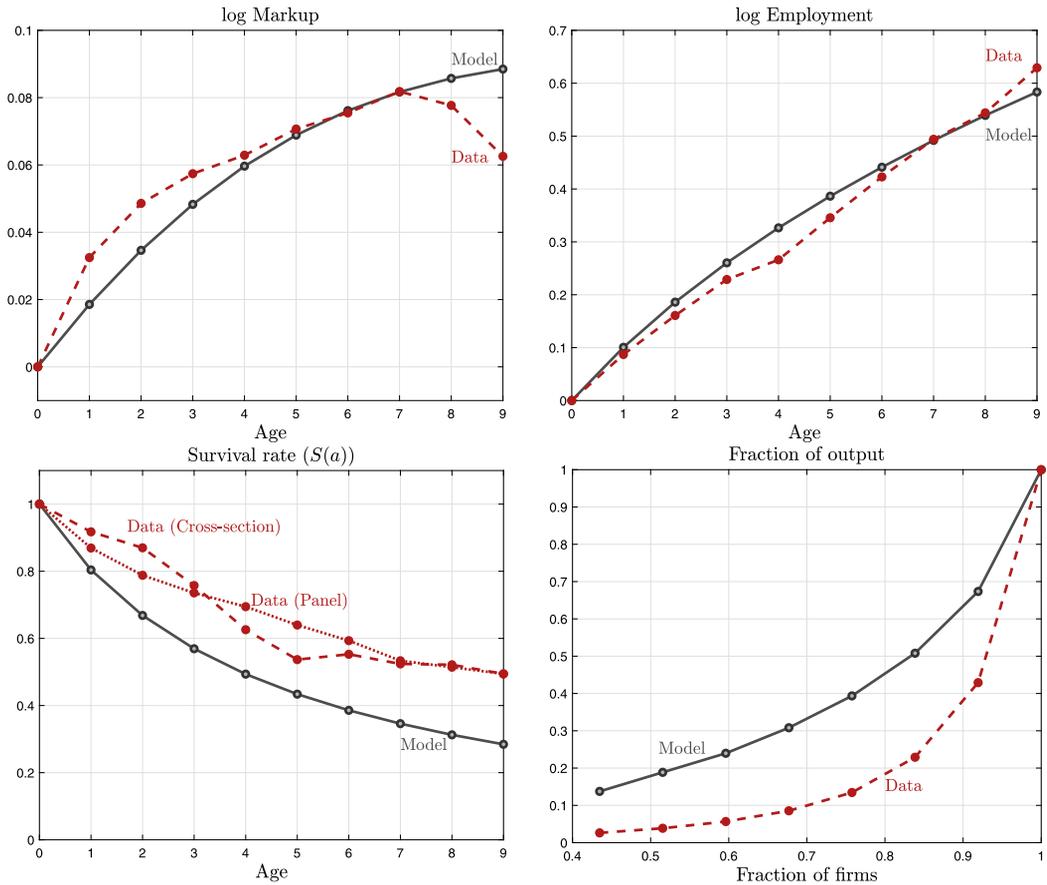

FIGURE 4.—NON-TARGETED MOMENTS: MODEL VERSUS DATA. *Notes*: The figure compares the model and the data for the life-cycle of markups (top-left panel), employment growth (top-right panel), survival (bottom-left panel), and the concentration of value added (bottom-right panel). Firm survival is measured both by the share of firms by age in the 2000 cross-section and by the share of firms of the entering cohort in 1991 by age. For the employment life-cycle, see Section A-2.3 in the Supplemental Material. To measure the concentration of value added, I drop the highest and lowest 3% of firms.

from the panel turn out to be quite similar. Second, the model captures the dynamics of exit relatively well, even though it slightly over-predicts the extent of "shake-out." The bottom-right panel displays the Lorenz curve of the firm size distribution, that is, the fraction of aggregate output that is accounted for by the smallest $x\%$ of firms. Both in the model and in the data, the curve is below the 45-degree line, reflecting the dispersion in firm size. The model, however, under-predicts the extent of concentration seen in the data.[22]

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-2.4 in the Supplemental Material, I derive an analytical expression for the average markup as a function of size. 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 model under-predicts the markup-size relationship for the following reason: the only reason 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. For example, the correlation between size and markups would be stronger if some firms expanded more efficiently and also had systematically higher quality draws, which would allow them to charge higher markups.

Finally, the model has implications for the sources of aggregate growth. Entrants account for $\frac{z}{\tau} \approx 13.5\%$ of aggregate creative destruction, and the share of aggregate growth accounted for by firms' own-innovation is about 70%. This is similar to Garcia-Macia, Hsieh, and Klenow (2019), who estimated that about 75%–80% of incumbent growth is due to own-quality improvements.[23] These two facts imply that the share of aggregate growth accounted for by entering firms is low at 4%. There are two main reasons why this is the case. First, 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. Second, I restrict the step size of innovation, λ, to be the same across all sources. If entrants were to enter with technologies that represented more drastic innovations, a given entry rate could be consistent with a larger entrant share of growth.

3.4. *Markups and Misallocation in Indonesia*

How large are the static efficiency losses due to heterogeneous markups? To see why answering this question in the context of a calibrated model is useful, note that the recent literature on misallocation stresses that resources are misallocated if revenue productivity

---

[22]In the model, firms' *only* margin of employment growth is to enter into new markets and to replace other producers; productivity growth in existing markets actually reduces employment by increasing markups. This sharp distinction is conceptually and analytically useful. It is, however, restrictive. For example, if the elasticity of demand exceeded unity (as in the theoretical extension described below), increases in quality would also lead 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, Hsieh, and Klenow (2019) and Luttmer (2010).

[23]Note, however, that own-quality improvements in Garcia-Macia, Hsieh, and Klenow (2019) increase employment (as markups are assumed to be constant), while such productivity increases in my model are fully reflected in firms' markups.

("TFPR") varies across producers (Restuccia and Rogerson (2008), Hsieh and Klenow (2009)). In my model, the variation in revenue productivity is entirely driven by variation in markups, because revenue productivity in product i is given by

$$\text{TFPR}_{it} \equiv \frac{p_{it}y_{it}}{w_t l_{it}} = \frac{p_{it}q_{it}}{w_t} = \mu_{it}. \tag{21}$$

Equation (21) highlights two reasons for the value of the theory. Most importantly, the markup μ_{it} is not an exogenous fundamental but is endogenously determined. Hence, the theory imposes structure on the part of the dispersion of measured TFPR, which can be attributed to variation in markups. Second, the aggregate statistics $\mathcal{M}$ and Λ depend on the distribution of markups 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 on the distribution of firm size. In environments where firms produce multiple products, the measured dispersion in *firm*-level markups will underestimate the welfare-relevant dispersion of markups across products.

The aggregate implications of market power are reported in Table V. The equilibrium churning intensity $\vartheta = \tau/I$ implied by Table III 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%, which is in the range of the estimates in De Loecker and Warzynski (2012), albeit at the lower end. Depending on the specification used, they estimated markups between 10% and 28%. They are also markedly lower than the results in De Loecker and Eeckhout (2017), who reported a sales-weighted average markup of 20% in 1980 and 60% in 2012 for publicly traded firms in the U.S. In my model, average markups are closely tied to the process of firm dynamics and a given aggregate growth rate. 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 across products of about 0.1. Because firms are active in multiple product markets, the dispersion across firms is lower. This discrepancy is seen in Figure 5, where I depict the equilibrium distribution of markups across products (dark bars) and firms (light bars). 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

TABLE V

MARKUPS AND MISALLOCATION IN INDONESIA[a]

	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

[a]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)).
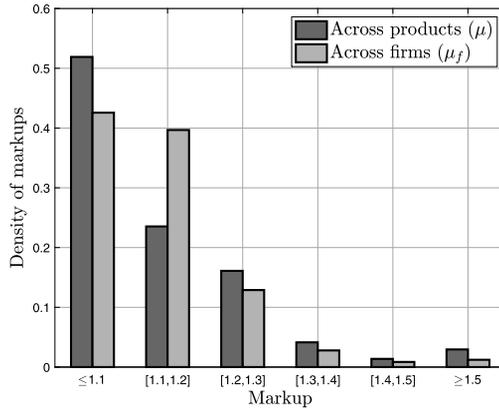
FIGURE 5.—THE STATIONARY DISTRIBUTION OF MARKUPS. *Notes*: The figure shows the distribution of markups at the product level (dark bars) and at the firm level (light bars). The results are based on the calibration reported in Table III.

by 10% relative to their social marginal product. This implied reduction in TFP seems small, especially compared to the much bigger numbers reported in Hsieh and Klenow (2009). There are two reasons. First, the model implies that markups account for only a small fraction of the observed dispersion in revenue products.[24] Empirically, the standard deviation of log labor shares is 0.74, which is consistent with Hsieh and Klenow (2009), who found numbers of around 0.7 in China and India. Hence, one of the reasons why the aggregate losses due to markups are relatively small is that the model attributes only a small share of the observed TFPR dispersion to markups. To explain a larger share, the extent of creative destruction had to be smaller. This, however, would be inconsistent with the observed rates of entry and life-cycle employment growth. The remainder could be explained by other frictions, adjustment costs, model misspecification, or measurement error.

Second, Hsieh and Klenow (2009) considered an elasticity of substitution of 3, whereas I impose a unitary demand elasticity. Recall that the change in aggregate TFP is approximately given by $d \ln \text{TFP} = -\frac{\sigma}{2} d \text{var}(\ln \text{TFPR})$, where σ is the elasticity of substitution across varieties. For $\sigma = 1$ and $d \text{var}(\ln \text{TFPR}) = 0.103^2$, one exactly recovers a TFP loss of 0.5% as reported in Table V.[25]

Stochastic Step Size and CES Preferences: Quantitative Implications

In Section 2.7, I generalized the theory by allowing for the step size to be stochastic and for a non-unitary demand elasticity. In this section, I show that these extensions 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. Consider first the model where the step size is drawn from $p_n = \frac{1-\kappa}{\kappa} \kappa^n$ (see Proposition 4). The left panel of Figure 6 shows the extent of misallo-

---

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

[25]The formula $d \ln \text{TFP} = -\frac{\sigma}{2} d \text{var}(\ln \text{TFPR})$, used in Hsieh and Klenow (2009), relies on the assumption that physical productivity and TFPR are independent and log-normally distributed. They estimated that $\text{var}(\ln \text{TFPR}^{\text{IND}}) - \text{var}(\ln \text{TFPR}^{\text{US}}) \approx 0.24$. For $\sigma = 3$, this implies a loss in aggregate TFP of 36%.
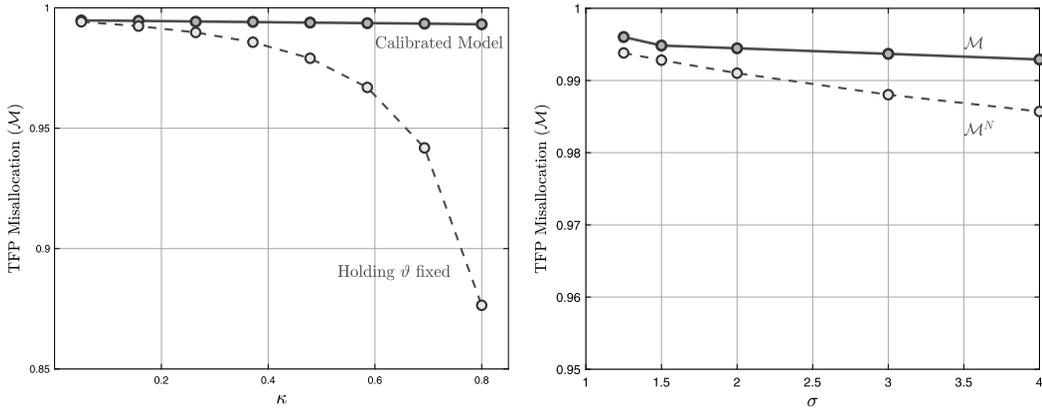
FIGURE 6.—MARKUPS AND MISALLOCATION: STOCHASTIC STEP SIZE AND CES PREFERENCES. *Notes*: The left panel displays the misallocation wedge $\mathcal{M}$ as a function of κ. The implications for the recalibrated model are shown as the solid line. For comparison, I also depict the case where I vary only κ but keep ϑ at its baseline value. The right panel shows the misallocation wedges $\mathcal{M}$ and $\mathcal{M}^N$ as a function of the demand elasticity σ. In the dark line, I show the expressions when the model is calibrated to the same moments as the baseline model. In the dashed line, I depict the "naive" measure, which keeps the distribution of markups at their baseline level.

cation $\mathcal{M}$ as a function of κ. The solid line corresponds to the case where—for every κ—I recalibrate all parameters to match the same moments as in my baseline calibration. The line is essentially flat and similar in magnitude to the baseline model. 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.

Recalibrating the model is crucial to reach this conclusion, because the parameter κ obviously has a mechanical effect on the distribution of markups, holding the churning intensity ϑ fixed. When I fix ϑ 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 churning does not respond endogenously.

The CES Model. The implied misallocation losses are also quantitatively robust to changes in the demand elasticity. Proposition 5 established that the misallocation wedge along the BGP is given by

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

This expression highlights that the demand elasticity σ has three effects. First, holding the distribution of quality gaps $\nu(\Delta)$ and the corresponding markups $\mu(\Delta)$ fixed, the demand elasticity σ determines how this heterogeneity is correctly aggregated. Second, σ directly determines the mapping from quality gaps Δ to the markups firms actually post. The higher σ, the lower the optimal CES markup $\frac{\sigma}{\sigma-1}$; that is, 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. Taking the distribution of markups from the baseline model as given (i.e., $G(\mu) = 1 - \mu^{-\theta}$, where θ stems from the baseline calibration), the misallocation wedge is given by

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

I refer to this measure as the "naive" measure, because it abstracts from any feedback from the demand elasticity to the markups firms actually post. This expression is nevertheless helpful, because $\mathcal{M}^N$ turns out to be a relatively tight upper bound for the losses from misallocation.

In the right panel of Figure 6, I depict this naive measure as a dashed line. 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 $\mathcal{M}$, which are depicted in the dark lines, are smaller. For a demand elasticity of 4, markups can reduce TFP by about a percentage point. Hence, the model with CES demand "adds" about half a percentage point of TFP losses relative to the baseline model. That $\mathcal{M}^N$ is 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, the endogenous distribution of quality gaps $\nu(\Delta)$ in the model with CES demand is such that low-markup products have more mass.

3.5. *The Importance of Expansion and Entry Costs*

The theory highlights that frictions for existing firms to expand into new product markets or costs for new firms to enter the economy are crucial determinants of creative destruction and hence of misallocation, aggregate growth, and the firm-size distribution. In the theory, such frictions are subsumed in the expansion and entry efficiency cost shifters φ_x and φ_z as they parameterize the resource requirements to introduce a technologically superior product to the market. Hence, they reflect not only the cost of innovation, but also institutional frictions such as license requirements or bureaucratic red tape.[26]

The theory implies that entry costs and expansion costs have distinct implications for the firm-size distribution, markups, and misallocation. Although both reduce creative destruction and increase misallocation, they affect the firm-size distribution differentially. Expansion costs reduce the share of creative destruction due to incumbents and lower average firm size. In contrast, higher entry costs lead to larger but fewer firms being active in equilibrium. This suggests that expansion costs might be of first-order importance in developing countries, because firms in these economies are mostly small and misallocation is argued to be rampant. On the other hand, the first-order effect of entry costs on important moments such as average firm size or the number of producers is counterfactual: high costs of entry go hand in hand with larger firms in equilibrium.

---

[26]One widely used measure of entry costs is developed in Djankov, La Porta, Lopez-De-Silvaes, and Shleifer (2002). Such variation in the regulation of entry has been linked to cross-country income differences in Barseghyan (2008), Bento (2016), and Herrendorf and Teixeira (2011).

To quantify the importance of these market barriers, I compare the calibrated model to data from the U.S. I target two moments that have a natural mapping to these barriers: the rate of entry and the extent of life-cycle employment growth. More specifically, I recalibrate the expansion and entry costs to match the life-cycle employment growth rate and the entry rate in the U.S. All remaining parameters are left unchanged. If these were indeed the only differences between the U.S. and Indonesia, the resulting estimates would be informative about the differences in misallocation and growth between these two economies. As an expositional shorthand, I therefore refer to the recalibrated economy as "the U.S."

For the rate of life-cycle employment growth in the U.S., I follow Hsieh and Klenow (2014), who reported that U.S. firms grow by a factor of 2 over a 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, Pugsley, and Sahin (2015), who reported a start-up rate between 8% and 11% for the whole economy, and Akcigit, Alp, and Peters (2020), who calculated an entry rate of 7.5% in the U.S. manufacturing sector.

The results of this exercise are contained in Table VI. 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 III. While both entry and expansion costs are lower in the U.S., the expansion margin is particularly important. The entry technology in the U.S. is about 15% more productive, and the costs for existing firms to break into new markets are about 33% 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, which calls for lower expansion costs in the U.S. 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, that is,

TABLE VI

COUNTERFACTUAL: CHANGES IN EXPANSION AND ENTRY COSTS[a]

	Heterogeneous Markups			Constant Markups
	Indonesia	U.S.	Change	Change
Calibration				
Entry rate	10.4%	8%		
Employment life-cycle	1.7	2		
Expansion costs	1	0.67	−33%	−21%
Entry costs	1	0.86	−14%	−2%
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, Misallocation, and Growth				
Life-cycle of markups	8.2%	6.2%	0.02	–
Average markup	11.73%	9.25%	−0.025	–
Markup dispersion (products)	10.5%	8.47%	−19.3%	–
Markup dispersion (firms)	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%	–
Growth rate (g)	3%	2.91%	−0.0009	−0.0004

[a]The first panel contains the calibration moments. The parameters for the Indonesian economy are contained in Table III. For the U.S. economy, I recalibrate the relative efficiency of expansion and entry, that is, φ_z and φ_x. All remaining parameters are the same as in Table III. In the last column, I report the results of a calibration of a model with constant markups.

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, a producer that is replaced in a particular product is less likely to exit. Entry barriers in the U.S. therefore also have to be lower to match the observed entry rate of 8%.

In the remaining panels, I report the equilibrium implications. I focus on two sets of results, which concern the distribution of firm size (Panel A) and the distribution of markups and aggregate growth (Panel B). Panel A shows that the estimated reductions in entry and expansion costs affect the firm-size distribution markedly. With firms having more opportunities to expand their scale of production, the equilibrium number of firms declines by 60% so that average firm size more than doubles. This reallocation comes especially at the expense of small firms, so that the output share of single-product firms declines by more than 70%. Hence, seemingly small differences in the rates of entry and life-cycle growth have large effects on the cross-sectional distribution of firm size.

Panel B shows that this shift toward large firms is accompanied by 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. While 7-year-old firms in Indonesia have about 8% higher markups than current entrants, this difference declines by two percentage points in the U.S.[27] At the aggregate level, average markups and their dispersion decline. 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 markups. 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 labor.

Finally, the last row contains the implications for aggregate growth. Interestingly, the economy-wide growth rate g hardly changes. This is due to the equilibrium effect on firms' own-innovation incentives. While creative destruction increases, firms' own-innovation efforts decline. That these different margins of growth are negatively related is not surprising; recall the optimality condition in (8), which showed that the marginal value of accumulating markups is discounted at rate $\rho + \tau$. In the calibrated model, this competition effect is sufficiently strong that $I + \tau$ is essentially constant, even though τ increases.

The results in Table VI have four implications. First, seemingly large changes in the stationary firm-size distribution and the number of active firms are fully consistent with small, empirically plausible differences in entry rates, employment life-cycle growth, and the increase in markups by age. Second, such large differences do not imply that countries grow at vastly different rates. The growth difference reported in Table VI accumulates to only a 2% productivity-level difference after 20 years. Third, 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. Whereas expansion frictions readily imply that many firms are small and experience little growth as they age, high entry costs would have the opposite implication.[28] Finally, while misallocation is indeed predicted to

---

[27]This is qualitatively consistent with the results reported in Hsieh and Klenow (2014), who argued that the increase in revenue-productivity by age, which in my model is proportional to markups, is steeper in India than in the U.S.

[28]In Section A-2.6 in the Supplemental Material, I use geographical variation across regions in Indonesia to report direct evidence for the importance of expansion barriers.

be higher in Indonesia, within the context of my model, the quantitative magnitude is relatively small, because 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, my model suggests that differences in markup heterogeneity are unlikely to account for a substantial share of it.

To analyze which of these implications are due to the presence of heterogeneous markups per se, in the last column of Table VI, I report the results of a calibration of a model with constant markups. This model, which I characterize in detail in the Supplemental Material, has the same structure as the model with CES preferences laid out in Section 2.7, except that I assume firms always charge a constant markup μ. I calibrate the exogenous markup μ to generate the same average markup as in the model with heterogeneous markups. Given that 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 own-innovation accounts for the same share of growth as in the model with heterogeneous markups (i.e., 70%). Although this model, of course, has no implications for the pattern of markups, Table VI shows that it also implies that frictions for existing firms are more important than entry costs, that firms are substantially smaller than in the U.S., 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. Hence, these conclusions do not depend directly on the presence of endogenous markups.

Robustness. In Section A-2.5 in the Supplemental Material, I examine the robustness of these results. First, I show that the results do not depend substantially on the choice of the curvature parameter ζ. I also study the sensitivity with respect to the underlying moments. The elasticity of average firm size and the share of small firms with respect to these two moments is sizable. If one were to, for example, assume that life-cycle employment growth in the U.S. was 2.5 instead of 2, the number of firms and the share of small firms would fall by 90%. By contrast, the effect on the growth rate is still almost indistinguishable from zero. An alternative calibration, which does not rely on the observed entry rate but explicitly exploits the size cutoff of census data, also delivers similar results.

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 are the result of a forward-looking, risky accumulation problem. Firms invest resources to increase the productivity of their existing products. Doing so allows them to increase 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 as well as 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 and product level. 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. This implies that markups are increasing

in age for the majority of firms. Furthermore, creative destruction is pro-competitive in that it reduces markup life-cycle growth and lowers markups, especially in the tail of the markup distribution.

As an application, I calibrate the model to firm-level panel data from Indonesia. Markups can plausibly account for 15% of the dispersion in measured revenue productivity 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 the main factor behind aggregate TFP differences across countries. I also find that firms in Indonesia are particularly hurt by frictions preventing existing firms from expanding into new product markets. By contrast, higher entry costs for new firms are less important. Quantitatively, such frictions increase 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, D., AND U. AKCIGIT (2012): "Intellectual Property Rights Policy, Competition and Innovation," *Journal of the European Economic Association*, 10, 1–42. [2039]

ACEMOGLU, D., U. AKCIGIT, H. ALP, N. BLOOM, AND W. KERR (2018): "Innovation, Reallocation, and Growth," *American Economic Review*, 108, 3450–3491. [2039,2060]

AGHION, P., AND P. HOWITT (1992): "A Model of Growth Through Creative Destruction," *Econometrica*, 60, 323–351. [2039,2043]

AGHION, P., A. BERGEAUD, T. BOPPART, P. J. KLENOW, AND H. LI (2019): "A Theory of Falling Growth and Rising Rents," Working Paper. [2039]

AKCIGIT, U., AND S. T. ATES (2019): "What Happened to US Business Dynamism?" NBER Working Paper 25756. [2039]

AKCIGIT, U., AND W. KERR (2018): "Growth Through Heterogeneous Innovations," *Journal of Political Economy*, 126, 1374–1443. [2039,2052]

AKCIGIT, U., H. ALP, AND M. PETERS (2020): "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," *American Economic Review* (forthcoming). [2039,2057,2068]

AMITI, M., AND J. KONINGS (2007): "Trade Liberalization, Intermediate Inputs and Productivity," *American Economic Review*, 97, 1611–1638. [2056]

ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2019): "The Elusive Pro-Competitive Effects of Trade," *Review of Economic Studies*, 86, 46–80. [2040]

ASKER, J., A. COLLARD-WEXLER, AND J. D. LOECKER (2014): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122, 1013–1063. [2040]

ATKESON, A., AND A. BURSTEIN (2008): "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 98, 1998–2031. [2040]

——— (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118, 433–484. [2054]

——— (2019): "Aggregate Implications of Innovation Policy," *Journal of Political Economy*, 127, 2625–2683. [2039]

AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. VAN REENEN (2017): "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 135, 645–709. [2039]

BARSEGHYAN, L. (2008): "Entry Costs and Cross-Country Differences in Productivity and Output," *Journal of Economic Growth*, 13, 145–167. [2067]

BARTELSMAN, E. J., J. C. HALTIWANGER, AND S. SCARPETTA (2013): "Cross-Country Differences in Productivity: The Role of Allocative Efficiency," *American Economic Review*, 103, 305–334. [2040]

BENTO, P. (2016): "Competition, Innovation, and the Number of Firms," Working Paper. [2067]

BENTO, P., AND D. RESTUCCIA (2017): "Misallocation, Establishment Size, and Productivity," *American Economic Journal: Macroeconomics*, 9, 267–303. [2040]

BERNARD, A., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93, 1268–1290. [2040,2048]

BERNARD, A. B., S. REDDING, AND P. SCHOTT (2010): "Multi-Product Firms and Product Switching," *American Economic Review*, 100, 70–97. [2056]

BUERA, F., AND R. N. FATTAL JAEF (2016): "The Dynamics of Development: Innovation, and Reallocation," Working Paper. [2040]

BUERA, F., J. KABOSKI, AND Y. SHIN (2011): "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101, 1964–2002. [2040]

DA-ROCHA, J.-M., M. M. TAVARES, AND D. RESTUCCIA (2017): "Policy Distortions and Aggregate Productivity With Endogenous Establishment-Level Productivity," NBER Working Paper 23339. [2040]

DAVID, J. M., AND V. VENKATESWARAN (2019): "The Sources of Capital Misallocation," *American Economic Review*, 109, 2531–2567. [2040,2065]

DAVID, J. M., H. A. HOPENHAYN, AND V. VENKATESWARAN (2016): "Information, Misallocation, and Aggregate Productivity," *Quarterly Journal of Economics*, 131, 943–1005. [2040]

DE BLAS, B., AND K. N. RUSS (2015): "Understanding Markups in the Open Economy," *American Economic Journal: Macroeconomics*, 7, 157–180. [2040,2048]

DE LOECKER, J. (2011): "Recovering Markups From Production Data," *International Journal of Industrial Organization*, 29, 350–355. [2057]

DE LOECKER, J., AND J. EECKHOUT (2017): "The Rise of Market Power and the Macroeconomic Implications," NBER Working Paper 23687. [2039,2064]

DE LOECKER, J., AND F. WARZYNSKI (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102, 2437–2471. [2039,2057,2059,2060,2064]

DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNIK (2016): "Prices, Markups and Trade Reform," *Econometrica*, 84, 445–510. [2039,2056,2057]

DECKER, R., J. HALTIWANGER, R. JARMIN, AND J. MIRANDA (2014): "The Role of Entrepreneurship in US Job Creation and Economic Dynamism," *Journal of Economic Perspectives*, 28, 3–24. [2039]

DHYNE, E., A. PETRIN, V. SMEETS, AND F. WARZYNSKI (2017): "Multi Product Firms, Import Competition, and the Evolution of Firm-Product Technical Efficiencies," NBER Working Paper 23637. [2056]

DJANKOV, S., R. LA PORTA, F. LOPEZ-DE-SILVAES, AND A. SHLEIFER (2002): "The Regulation of Entry," *Quarterly Journal of Economics*, 117, 1–37. [2067]

EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2015): "Competition, Markups and the Gains From International Trade," *American Economic Review*, 105, 3183–3221. [2040]

——— (2018): "How Costly Are Markups?" Working Paper. [2039]

EPIFANI, P., AND G. GANCIA (2011): "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 83, 1–13. [2040]

FATTAL JAEF, R. N. (2018): "Entry and Exit, Multi-Product Firms, and Allocative Distortions," *American Economic Journal: Macroeconomics*, 10, 86–112. [2040]

FOSTER, L., J. C. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98, 394–425. [2039]

GABAIX, X. (1999): "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 114, 739–767. [2048]

GARCIA-MACIA, D., C.-T. HSIEH, AND P. KLENOW (2019): "How Destructive Is Innovation?" *Econometrica*, 87, 1507–1541. [2039,2063]

GROSSMAN, G., AND E. HELPMAN (1991): "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 58, 43–61. [2039,2043]

HALTIWANGER, J., R. DECKER, AND R. JARMIN (2015): "Top Ten Signs of Declining Business Dynamism and Entrepreneurship in the US," in *Kauffman Foundation New Entrepreneurial Growth Conference*. [2039]

HERRENDORF, B., AND A. TEIXEIRA (2011): "Barriers to Entry and Development," *International Economic Review*, 52, 573–602. [2067]

HOLMES, T. J., W.-T. HSU, AND S. LEE (2014): "Allocative Efficiency, Mark-Ups, and the Welfare Gains From Trade," *Journal of International Economics*, 92, 195–206. [2040]

HOPENHAYN, H. (2014): "On the Measure of Distortions," Working Paper. [2040]

HSIEH, C.-T., AND P. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 771–807. [2038,2040,2055,2064,2065]

——— (2014): "The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*, 129, 1035–1084. [2040,2068,2069]

HSIEH, C.-T., AND B. A. OLKEN (2014): "The Missing 'Missing Middle'," *Journal of Economic Perspectives*, 28, 89–108. [2056]

HURST, E., AND B. PUGSLEY (2012): "What Do Small Businesses Do?" *Brookings Papers on Economic Activity*, Fall, 73–142. [2057]

JONES, C., AND J. KIM (2018): "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, 126, 1785–1826. [2049]

KARAHAN, F., B. PUGSLEY, AND A. SAHIN (2015): "Understanding the 30-Year Decline in the Start-Up Rate: A General Equilibrium Approach," Working Paper. [2068]

KLETTE, T. J., AND S. KORTUM (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112, 986–1018. [2037,2039,2041,2043,2046,2051,2052,2059]

LA PORTA, R., AND A. SHLEIFER (2009): "The Unofficial Economy and Economic Development," *Brookings Papers on Economic Activity*, Fall, 275–352. [2057]

——— (2014): "Informality and Development," *Journal of Economic Perspectives* 28, 109–126. [2057]

LENTZ, R., AND D. T. MORTENSEN (2008): "An Empirical Model of Growth Through Product Innovation," *Econometrica*, 76, 1317–1373. [2039]

LUTTMER, E. G. J. (2010): "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, 2, 547–576. [2063]

MELITZ, M., AND G. OTTAVIANO (2008): "Market Size, Trade and Productivity," *Review of Economic Studies*, 75, 295–316. [2040]

MIDRIGAN, V., AND D. Y. XU (2014): "Finance and Misallocation: Evidence From Plant-Level Data," *American Economic Review*, 104, 422–458. [2040]

MOLL, B. (2014): "Productivity Losses From Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *American Economic Review*, 104, 3186–3221. [2040]

PETERS, M. (2020): "Supplement to 'Heterogeneous Markups, Growth, and Endogenous Misallocation'," *Econometrica Supplemental Material*, 88, https://doi.org/10.3982/ECTA15565. [2040]

PETERS, M., AND C. WALSH (2019): "Population Growth and Firm Dynamics," Working Paper. [2039]

RESTUCCIA, D., AND R. ROGERSON (2008): "Policy Distortions and Aggregate Productivity With Heterogeneous Establishments," *Review of Economic Dynamics*, 11, 707–720. [2038,2040,2064]

SCHOAR, A. (2010): "The Divide Between Subsistence and Transformational Entrepreneurship," *Innovation Policy and the Economy*, 10, 57–81. [2057]