Heterogeneous Markups, Growth and Endogenous Misallocation

Michael Peters*
Yale University
November 2018

Abstract

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

*eMail: m.peters@yale.edu. I am especially grateful to Daron Acemoglu, Abhijit Banerjee and Rob Townsend for their invaluable guidance. I also thank Ufuk Akcigit, Susanto Basu, Penny Goldberg, Pete Klenow, Sam Kortum, Giuseppe Moscarini, Ezra Oberfield and Aleh Tsivinski as well as Gianluca Violante and four anonymous referees, whose suggestions benefited the paper substantially. I am also grateful for the financial assistance of the George and Obie Shultz fund to acquire the data necessary for this project.
1 Introduction

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

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

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

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

Despite its parsimony, the model has rich testable implications for the empirical patterns of markups, size and age at the firm-level. The model predicts markups to follow a distinct life cycle pattern within products: conditional on survival, markups increase stochastically as the result of the firm’s productivity-enhancing investments. To aggregate these product-level dynamics to the firm-level, two counteracting effects are at play. On the one hand, firms increase markups in their existing products as they age. This “own-innovation channel” raises the average markup of old firms relative to young firms. On the other hand, firms also expand into new products and lose existing products to other firms. This “creative-destruction channel” lowers the average markup of old firms as markups in new products are on average lower than in the old, infra-marginal products the firm loses. The “own-innovation channel” therefore increases the elasticity of markups by age and reduces the extent to which firms grow...
in size as they get older. The “creative-destruction channel” has the exact opposite effect: it increases the rate of life cycle employment growth at the expense of profitability. I derive an explicit formula for the life cycle of markups at the firm-level, which shows 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. Furthermore, the slope of the markup life cycle is smaller, the more important the “creative-destruction channel”.

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

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

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

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

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

A growing empirical literature shows that markups vary systematically in the cross-section of firms and respond to changes in the environment. In particular, markups are low for entering firms (Foster et al. (2008)), high for exporters (De Loecker and Warzynski, 2012), increase in response to trade liberalizations (De Loecker et al., 2016) and are argued to be an important source of variation in revenue-based productivity measures (De Loecker, 2011a). Recently, there has also been a growing interest in the aggregate implications of market power. De Loecker and Eeckhout (2017) and Autor et al. (2017) for example argue that markups have been rising and that high profitability, “superstar” firms, have become more important for the aggregate economy. The theory proposed in this paper is qualitatively consistent with these facts as it predicts that markups are increasing in age and size. Furthermore, the theory suggests that markups and concentration will rise in the response to a decline in churning. Coincidentally, some authors have indeed argued that dynamism has been declining (Decker et al., 2014; Haltiwanger et al., 2015).

The application in this paper is closely related to the literature on misallocation (see e.g. Hsieh and Klenow (2009), Restuccia and Rogerson (2008), Bartelsman et al. (2013) and the survey article by Hopenhayn (2012)). This literature typically treats misallocation as an exogenous firm-specific wedge. The model proposed in this paper generates misallocation endogenously. While numerous theories of misallocation based on inefficient input use have been proposed (e.g. imperfect capital markets (Buer et al., 2011; Moll, 2014; Banerjee and Moll, 2010; Midrigan and Xu, 2014), information frictions (David et al., 2016) or adjustment costs Asker et al. (2014)), this paper is the first to quantitatively explore the role of monopolistic power as a source of misallocation.

A recent literature has also studied how exogenous distortions affect the patterns of firm growth and entry (see e.g. Hsieh and Klenow (2014), Bento and Restuccia (2017), Da-Rocha et al. (2017), Fattal Jaef (2018) and Buera and Jaef (2016)). In this paper, I take the opposite approach as misallocation is fully endogenous and hence emerges as an equilibrium outcome together with the size distribution of firms. This allows me to study how particular policies affect both misallocation and firm size. I show that barriers for incumbent firms to enter new product markets increase misallocation and makes firms small. This is qualitatively consistent with the cross-country evidence (see e.g. Hsieh and Olken (2014) or Poschke (2018)). In contrast, entry costs also induce misallocation but increase firm size. The difference between frictions for existing firms to enter new product markets and entry costs for new producers is also discussed in Bento (2016) in a static environment.

While this is, to the best of my knowledge, the first paper that focuses on imperfect output markets in relation to the literature on growth and misallocation, there is a growing literature in the field of international trade stressing the importance of markups. On the theory side, Bernard et al. (2003), Melitz and Ottaviano (2008) or Atkeson and Burstein (2008) are examples of theories that generate heterogeneous markups. That the welfare gains from trade are affected by markup heterogeneity and misallocation is explicitly analyzed in Edmond et al. (2015), Epifani
and Gancia (2011), Arkolakis et al. (2016), De Blas and Russ (2015) and Holmes et al. (2014). In a recent paper, Edmond et al. (2018) analyze the allocative consequences of markup dispersion in an industry equilibrium model with free entry. In contrast to the model of this paper, all these frameworks assume that firm efficiency is exogenous.

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

2 Theory

2.1 The Environment

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

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

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

$$\ln Y_t = \int_{0}^{1} \ln \left( \sum_{i \in S_t} y_{fit} \right) \, di,$$

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

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

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

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

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

Consider first the static allocations for a given number of firms and distributions of productivity \( q \). Given that production takes place with a constant returns to scale technology, firms compete in prices and different brands of variety \( i \) are perceived as perfect substitutes, in equilibrium only the most productive firm within a market \( i \) will be active. However, the presence of competing producers (even though they are less efficient) imposes a constraint on the leading firm’s price setting. Because the demand function associated with (1) has a unitary price elasticity, the most efficient firm will resort to limit pricing. Letting \( q_{it} \) denote the productivity of the actual producer, i.e. the most efficient firm within the market, the equilibrium markup in market \( i \) is given by

\[
\mu_{it} \equiv \frac{p_{it}}{w_{it}/q_{it}} = \frac{w_{it}/q_{it}^{F}}{w_{it}/q_{it}} = \frac{q_{it}}{q_{it}^{F}},
\]

where \( w_{it} \) denotes the equilibrium wage and \( w_{it}/q_{it}^{F} \) is the marginal cost of the second most productive firm, which I refer to as the follower. Intuitively, a bigger productivity advantage shields the current producer from competition and allows him to post a higher markup.

From (2) one can derive the allocation of labor at the firm-level. The Cobb-Douglas assumption in (1) implies that sales are equalized across markets, i.e. \( p_{it}y_{ift} = Y_{t} \). Letting \( N_{ft} \) be the set of markets firm \( f \) is active in, total employment of firm \( f \), \( l_{ft} \), is given by

\[
l_{ft} = \sum_{i \in N_{ft}} l_{fit} = \sum_{i \in N_{ft}} \frac{1}{q_{it}^{F}} q_{it} y_{ift} = \sum_{i \in N_{ft}} \frac{1}{q_{it}^{F}} q_{it} \frac{Y_{t}}{w_{it}} = \frac{Y_{t} n_{ft}}{w_{it}} \times \left( \frac{1}{n_{ft}} \sum_{i \in N_{ft}} \mu_{it}^{-1} \right) = \frac{Y_{t} n_{ft}}{w_{it}} \times \mu_{f}^{-1},
\]

where \( n_{ft} = |N_{ft}| \) is the number of markets catered to by firm \( f \) and the last equality defines the average markup at the firm-level as

\[
\mu_{f} = \left( \frac{1}{n_{ft}} \sum_{i \in N_{ft}} \mu_{it}^{-1} \right)^{-1}.
\]

Equation (3) highlights that the size of a firm is shaped by two forces. Firms are large if they are active in many markets \( n_{ft} \), i.e. expanding into novel markets increases employment at the firm-level. Conversely, for a given number of markets \( n_{ft} \), higher markups reduce firm-employment, so that variation in markups induces variation in employment holding the number of markets fixed.

Given the above structure, the economy has a transparent aggregate representation. Letting \( L_{Pt} \) denote the total mass of production workers, (3) implies that

\[
L_{Pt} = \int_{f} l_{ft} df = \frac{Y_{t}}{w_{t}} \int_{f} \sum_{i \in N_{ft}} \mu_{it}^{-1} df = \frac{Y_{t}}{w_{t}} \times \left( \int_{0}^{1} \mu_{it}^{-1} di \right).
\]

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

\[
w_{t} = \exp \left( \int_{0}^{1} \ln q_{it}^{F} di \right) = \exp \left( \int_{0}^{1} \ln q_{it} di \right) = Q_{t} \times \exp \left( \int_{0}^{1} \ln \mu_{it}^{-1} di \right),
\]

\(^{1}\)That markups are fully determined from limit pricing makes the dynamic decision problem of the firm very tractable. In Section OA-1.3 in the Online Appendix I discuss more specifically why the model would be far less tractable if firms were assumed to compete a la Cournot instead of Bertrand.
where \( q_t = \int_0^1 \ln q_{it} \, di \) is the usual CES efficiency index. Aggregate output is therefore given by

\[
Y_t = Q_t M_t L_t P_t \quad \text{where} \quad M_t = \exp \left( \int_0^1 \ln \mu_{it}^{-1} \, di \right) = \exp \left( \frac{E \left[ \ln \mu_{it}^{-1} \right]}{E \left[ \mu_{it}^{-1} \right]} \right). \tag{4}
\]

Equation (4) highlights the allocative consequences of market power. Aggregate TFP is the product of firms’ physical productivity measure \( Q_t \) and the term \( M_t \), which summarizes the degree of markup misallocation. In the absence of markups, i.e. \( \mu_{it} = 1 \), it follows that \( M_t = 1 \). Moreover, (4) implies that \( M_t \leq 1 \) and that \( M_t = 1 \) if and only if markups are equalized. Hence, aggregate TFP depends on the dispersion of markups. While a common proportional increase in markups leaves the degree of misallocation unchanged, higher markup dispersion reduces allocative efficiency and hence aggregate TFP.

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

\[
\Lambda_t = \frac{w_t L_{P,t}}{Y_t} = \left( \int_0^1 \mu_{it}^{-1} \, di \right) = E \left[ \mu_{it}^{-1} \right]. \tag{5}
\]

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

Equations (4) and (5) highlight that the static macroeconomic implications of market power are fully summarized by two aggregate statistics \( M_t \) and \( \Lambda_t \). In the terminology of Chari et al. (2007), \( \Lambda_t \) and \( M_t \) are akin to a labor wedge and an efficiency wedge. Moreover, both statistics only depend on the marginal distribution of markups.\(^2\)

Below I construct this marginal distribution as an endogenous outcome from firms’ innovation decisions so that misallocation, factor shares and growth are jointly determined.

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

\[
TFPR_{it} = \frac{p_{it} y_{it}}{w_{it} L_{P,t}} = \frac{p_{it} q_{it}}{w_{it}} = \mu_{it}. \tag{6}
\]

Hence, revenue productivity is equalized if and only if producers post a common markup. In the frameworks of Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), revenue productivity is determined by exogenous distortions (“firm-specific wedges”).\(^3\) Equation (6) highlights that a firm charging a high markup has high revenue productivity and hence would be identified as facing a high distortionary tax. Intuitively: market power makes

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

\(^3\)In these contributions TFPR is proportional to \( (1 + \tau_{K,i})^\alpha \), where \( \tau_{K,i} \) and \( \tau_{Y,i} \) are exogenous firm-specific wedges on capital and output and \( \alpha \) is the capital share.
firms too small. However, the firm might not be constrained or subject to policy frictions but rather chose to under-produce by setting a high price.

Equation (6) also highlights two aspects of my theory, which are absent from the above mentioned literature on misallocation. Most importantly, the markup $\mu_{it}$ is not an exogenous fundamental but is endogenously determined with the evolution of firms physical efficiencies. Secondly, the aggregate statistics $M_t$ and $\Lambda_t$ depend on the distribution of markup across products. Empirically, revenue productivity is usually measured at the firm-level. The mapping between the macroeconomic consequences of misallocation and the firm-level data therefore depends crucially on the distribution of firm size. In particular, in environments where firms produce multiple products, the measured dispersion in firm-level markups will underestimate the dispersion of markups across products, which is welfare-relevant.

2.3 Dynamics: Innovation and Creative Destruction

Both the production possibility frontier (as summarized by $Q_t$) 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 $\lambda > 1$.\(^4\) This structure is convenient because it implies that equilibrium markups are given by (see (2))

$$
\mu_{it} = \frac{q_{it}}{q_i^F} = \frac{\lambda^{r_{it}}}{\lambda^{r_{it}^F}} = \lambda^{r_{it}^F - r_{it}} \equiv \lambda^{\Delta_{it}}, \quad (7)
$$

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

In a given market $i$, productivity increases can stem from three distinct sources: (i) a new firm can enter market $i$ with a new technology, (ii) an existing firm, who is not currently active in market $i$, can expand into this market and (iii) the current producer in market $i$ can increase his productivity to gain additional monopoly power. I assume that these three sources of growth are fully symmetric in how they improve the current frontier technology: if the current productivity in market $i$ is given by $q_{it}$, the new productivity is given by $\lambda q_{it}$.

Treating productivity increases through current and new producers symmetrically is not only standard in most Schumpeterian models of growth, but is particularly appealing in the current context, as it highlights the different allocative consequences of creative destruction and own-innovation. While new producers and incumbents increase the frontier technology by the same amount, the implications for equilibrium markups are very different. In particular, equation (7) implies that different sources of productivity growth have different implications for the evolution of markups. In case the innovation stems from the current producer in market $i$, the equilibrium markup increases by a factor $\lambda$, as the quality gap rises from $\Delta$ to $\Delta + 1$. In contrast, when productivity growth is due to creative destruction, whereby an entirely new firm or an existing firm enters market $i$ as a novel producer, the equilibrium markup decreases by a factor $\lambda^{\Delta_{(i,t)} - 1}$, as the new producer is only a single step ahead on the quality ladder, i.e. the quality gap declines from $\Delta$ to unity.\(^5\)

Given this structure, the state of the firm is in principle a multi-dimensional object: the number of products $n$, the quality of each of these products $[q_{j}]_{j=1}^{n}$ and the quality-gaps in each product line $[\Delta_{j}]_{j=1}^{n}$. However, the

\(^4\)Specifically, letting $r$ denote the rung of the ladder, qualities are ordered according to $q_{r+1} = \lambda q_r$.

\(^5\)In Section 2.9 I generalize the analysis to a setting where the step size is not necessarily equal to unity but drawn from a distribution. Note also that the continuous time formulation of the model precludes the possibility that a variety experiences both entry and a productivity improvement by the current producer.
Cobb-Douglas demand structure implies that equilibrium profits in market $i$ are given by

$$\pi_{it} = (1 - \mu_{it}^{-1}) Y_t = (1 - \lambda^{-\Delta i}) Y_t \equiv \pi_i (\Delta_i), \quad (8)$$

i.e. they only depend on the quality gap and not on the level of quality $q_{it}$. I therefore restrict attention to equilibria where firm behavior only depends on the payoff-relevant state variables $\left( n, [\Delta_i]_{j=1}^n \right)$.

Firms can spend resources to improve the productivity in the markets they are currently producing in and can try to break into novel markets. I adopt the usual stochastic formulation, whereby firms can choose the flow rates of increasing the productivity of existing products and of expanding in a novel, randomly selected market. I denote the rate of own-innovation on existing products by $[I_{it}]_{i=1}^n$ and the rate of expansion by $[x_{it}]_{i=1}^n$. The associated cost function (denoted in units of labor) is given by $\Gamma \left( [x_i, I_i]_{i=1}^n ; n, [\Delta_i]_{j=1}^n \right)$.

Optimal behavior is then described by the value function $V_t (n, [\Delta_i])$, which is given by

$$r_t V_t (n, [\Delta_i]) - \dot{V}_t (n, [\Delta_i]) = \sum_{i=1}^n \pi_t (\Delta_i) - \sum_{i=1}^n \tau_t \left[ V_t (n, [\Delta_i]) - V_t (n - 1, [\Delta_i]_{j \neq i}) \right] + \left\{ \begin{array}{ll} \max_{[x_i, I_i]_{i=1}^n} & \sum_{i=1}^n I_i \left[ V_t \left( n, \left[ \Delta_i \right]_{j \neq i}, \Delta_i + 1 \right) \right] - V_t (n, [\Delta_i]) \right. \\
& + \left. \sum_{i=1}^n x_i \left[ V_t \left( n + 1, \left[ \Delta_i, 1 \right] \right) - V_t (n, [\Delta_i]) \right] - \Gamma \left( [x_i, I_i]; n, [\Delta_i] \right) w_t \right\} \quad (9)$$

The value of the firm consists of three parts. First there is the total flow payoff, which is simply the sum of profits across all markets. Second, there is the possibility of losing any of the $n$ existing products to other firms. This happens at the endogenous rate of creative destruction $\tau$, which is determined in equilibrium. Third, there are two option values of innovation as the firm has the option to increase the markup in each of its markets through own-innovation $I_i$ and the firm can spend resources to expand into new markets with flow rate $x_i$. Note that the quality gap in novel markets is always equal to unity, i.e. new products have low markups. Finally, the last term in (9) captures the cost of own-innovation and market expansion.

While the recursive formulation for $V_t$ in (9) looks daunting, it turns out that $V_t$ admits a simple closed form solution. I assume a particular, additively separable functional form for the cost function $\Gamma (\cdot)$, which is consistent with balanced growth and allows me to derive an analytic solution

$$\Gamma \left( [x_i, I_i]; n, [\Delta_i] \right) = \sum_{i=1}^n c \left( I_i, x_i; \Delta_i \right) \text{ where } c \left( I, x; \Delta \right) = \lambda^{-\Delta} \frac{1}{\varphi_I I^\zeta} + \frac{1}{\varphi_x x^\zeta}. \quad (10)$$

Here $\varphi_I$ and $\varphi_x$ parametrize the efficiency of the innovation and expansion technology and $\zeta > 1$ ensures that the cost function is convex so that there is a unique solution. The cost shifters $\varphi_I$ and $\varphi_x$ comprise both technological features of the innovation process as well as institutional determinants like bureaucratic requirements to produce

---

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

8
a particular product. In addition, both cost functions contain scaling variables, which make the model consistent with balanced growth (Sutton, 1997; Luttmer, 2010).

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

$$V_t (1, 1) \leq \frac{1}{\varphi_z} w_t = 0 \text{ with equality if } z > 0.$$  \hfill (11)

For the remainder of the paper, I focus on the case with positive entry, where the condition in (11) holds with equality. The aggregate rate of creative destruction $\tau_t$ is therefore given by

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

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

### 2.4 The Stationary Equilibrium

Given this set-up, I now characterize the stationary (or balanced-growth-path) equilibrium of this economy. Labor market clearing requires that production labor $L_{Pt}$ and research labor $L_{Rt}$ add up to the aggregate labor endowment, which I normalize to unity. Hence, letting $z_t$ be the aggregate entry rate and $x_{it}$ and $I_{it}$ the innovation flow rates chosen of the producer of product $i$, (10) and (11) imply that

$$1 = L_{Pt} + L_{Rt} = L_{Pt} + \frac{1}{\varphi_z} z_t + \int_{i=0}^{1} \left( \frac{1}{\varphi_x} x_{it}^\zeta + \lambda - \Delta \right) I_{it}^\zeta di.$$  \hfill (12)

A stationary equilibrium is then defined in the usual way.

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

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

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

---

7The term $\lambda \Delta$ in $c^\Delta (.)$ implies that innovations are easier the bigger the within-market productivity advantage $\Delta$. This is similar in spirit to the assumption of knowledge capital made in Klette and Kortum (2004) or the setup in Atkeson and Burstein (2010). Intuitively: per-period profits are given by $(1 - \lambda \Delta) Y$ (see (8)) and hence concave in $\Delta$. For innovation incentives to be constant, the marginal costs of innovation have to be lower for more advanced firms. The leading term in (10) $(\lambda \Delta)$ is exactly the right normalization to balance those effects. Note that firms only generate a high productivity gap when they have multiple innovation in a row. Hence, (10) effectively posits that firms can build on their own innovations of the past. The term $n^\zeta$ in $c^\Delta (.)$ serves a similar purpose and implies that the cost of expanding at rate $x$ per market (i.e. $nx = X$) is linear in $n$. 

1. The value function is given by

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

where

\[ V_t^P = \frac{\pi (1) + (\zeta - 1) \frac{\phi_x}{\varphi_x} w_t}{\rho + \tau} \quad \text{and} \quad V_t^M(\Delta) = \frac{\pi (\Delta_i) - \pi (1) + (\zeta - 1) \lambda - \Delta (i) \varphi_x w_t}{\rho + \tau}, \]

2. The optimal rates of innovation, expansion, entry and creative destruction, \((I_t, x_t, z_t, \tau_t)\), are constant and given by \((I, x, z, \tau)\). In particular, the optimal expansion rate \(x\) is given by

\[ x = \left( \frac{\phi_x}{\phi_z} \right)^{\frac{1}{\zeta - 1}}, \]

and the innovation rate \(I\) solves

\[ I = \left( \frac{\lambda - 1}{\lambda} \frac{1}{\rho + \tau} \left( \frac{\phi_x}{\xi} \frac{Y_t}{w_t} - \frac{\zeta - 1}{\zeta} I \right) \right)^{\frac{1}{\zeta - 1}}, \]

3. The distribution of markups is stationary so that the equilibrium wedges \(M\) and \(\Lambda\) are constant,

4. All aggregate variables grow at the common growth rate

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

**Proof.** See Section A-1.1 in the Appendix. The condition \(\rho > \frac{\zeta - 1}{\zeta} \left( \frac{1}{\varphi_x} \right)^{1/(\zeta - 1)}\) ensures that the free entry condition is satisfied.

Proposition 1 establishes that the economy permits a unique stationary equilibrium, which can essentially be characterized analytically. The value of the firm, \(V_t(n, [\Delta_i])\), has an intuitive structure. First of all, it is additive across products. Secondly, the value of producing a given product with quality gap \(\Delta_i\) also consists of two additive parts. The first term, \(V_t^P\), captures the value of producing a particular product with a quality gap of unity (and hence a markup of \(\lambda\)). It consists of the production value and the inframarginal rents of the concave expansion technology, i.e. the option value of being able to expand in new markets. This part of a firm’s value scales linearly in the number of markets \(n\) and is similar to the baseline model of Klette and Kortum (2004). The second term, \(V_t^M(\Delta)\), captures the possibility of accumulating market-power. It consists of the flow value of being able to charge higher markups, i.e. \(\pi (\Delta_i) - \pi (1)\), augmented by the possibility of increasing markups even further. 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 \(\rho\), but also the rate of creative destruction \(\tau\) to account for the risk of being replaced.

Associated with the value function \(V_t\) are optimal innovation, entry and expansion decisions, which are constant - both across products and across time. Equation (14) shows that the rate at which incumbents enter new markets has a simple closed form expression and depends on the efficiency of incumbent creative destruction \((\phi_x)\) relative to the one of entrants \((\phi_z)\). The optimal extent of own-innovation \(I\) depends on two endogenous aggregate variables - the rate of creative destruction \(\tau\) and size of the market \(\frac{Y_t}{w_t}\) (relative to the cost of innovation). An increase in the rate of creative destruction \(\tau\) reduces firms’ incentives to engage in own-innovation, as the expected time-horizon of
the accrual of monopolistic rents becomes shorter. Conversely, innovation incentives are high if aggregate demand \(Y_t\) is large relative to the cost of innovation \(w_t\).

The equilibrium entry rate \(z\) is then determined from the labor market clearing condition (12). The rate of creative destruction is therefore given by \(\tau = z + x\). The aggregate rate of growth is simply given by the rate of growth of technology \(Q_t\), because the efficiency wedge \(\Lambda_t\) is constant in a stationary equilibrium. Because all three sources of innovation generate productivity improvements of the same size, \(g\) is proportional to the sum of creative destruction \(\tau\) and firms’ rate of own-innovation \(I\).

### 2.5 The Cross-Sectional Distributions of Markups

The main theoretical contribution of this paper is to derive the equilibrium distribution of markups and hence the endogenous degree of misallocation from firms’ entry, expansion and innovation policies. In a stationary equilibrium, the distribution of markups is time-invariant and can be characterized explicitly. To construct this distribution, recall that markups only depend on the distribution of quality gaps \(\Delta\) across markets (see (7)). The cross-sectional distribution of markups is therefore fully characterized by \(\{\nu(\Delta, t)\}_{\Delta=1}^{\infty}\), where \(\nu(\Delta, t)\) denotes the measure of markets with quality gap \(\Delta\) at time \(t\). These measures solve the set of differential equations

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

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

Equation (16) is the key equation to characterize the equilibrium distribution of markups. Three properties are noteworthy. First of all, the distribution is fully determined from the two endogenous variables \((I, \tau)\) and is hence jointly determined with the economy-wide growth rate \(g\). Secondly, the distribution of firm size is not required to solve for the distribution of markups across products. This is due to the fact that all firms innovate and expand at constant rates \(I\) and \(x\) per market. Finally, (16) highlights the pro-competitive effects of creative destruction: while productivity growth by existing producers are markup increasing, market churning through creative destruction shifts the distribution of markups downwards. This suggests that creative destruction is a force, which tends to reduce misallocation. Firms’ own-innovation efforts in contrast lower allocative efficiency through higher and more dispersed markups. The next proposition shows that this intuition is exactly correct.

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

\[
\vartheta_I = \frac{\tau}{I} \text{ and } \theta = \frac{\ln(1 + \vartheta_I)}{\ln \lambda}.
\]

---

8More specifically, together with the free entry condition (11) and the equilibrium labor wedge \(\Lambda_t\) in (5), (12) and the three relationships (13), (14) and (15) are six equations in the six unknowns \((x, \tau, I, \frac{\nu}{w}, \frac{\lambda}{w}, L, P_t)\). In Section A-1.1 I show how these can be reduced to a system of two equations in two unknowns.

9Recall that the product space has measure one. Because the aggregate rate of entry is given by \(z\), this is also the rate at which each product is subject to entry by a new firm. Similarly, because each existing firm innovates at rate \(x\) per product, the aggregate rate at which existing firms expand into the new markets is also given by \(x\).
Then:

1. **The distribution of markups is given by**
   \[
   G(\mu) = 1 - \mu^{-\theta},
   \]

2. **The degree of misallocation \( M \) and the labor wedge \( \Lambda \) are given by**
   \[
   M = e^{-1/\theta} \frac{1 + \theta}{\theta} \quad \text{and} \quad \Lambda = \frac{\theta}{1 + \theta}. 
   \] 

(17)

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

Proposition 2 contains the main theoretical result of this paper: the cross-sectional distribution of markups \( G(\mu) \), the extent of misallocation \( M \) and the labor wedge \( \Lambda \) are jointly determined with firms’ innovation incentives and the rate of creative destruction. In particular, the endogenous distribution of markups takes a pareto form, whose shape parameter \( \theta \) is endogenous and fully determined from a single endogenous statistic - the **churning intensity** \( \vartheta_I \). This statistic measures the speed with which firms are being replaced by new producers relative to firms’ own-innovation efforts. If churning is intense, the shape parameter is large so that both markup heterogeneity and the average markup decline. If in contrast churning is of little importance, the resulting distribution of markups has a fat tail and both the average markup and their dispersion is large. Because firms’ own-innovation incentives \( I \) and the rate of creative destruction \( \tau \) are determined endogenously and hence depend on parameters or policies, markups and misallocation will also endogenous respond. This is the crucial difference to Bernard et al. (2003), who generate a Pareto distribution of markups from firms’ exogenous productivity draws.

The macroeconomic consequences of this endogenous markup distribution are fully summarized the two sufficient statistics \( M \) and \( \Lambda \), which naturally also only depend on \( \vartheta_I \) and have the closed-form representation given in (17). In particular, it is easy to verify that both \( M \) and \( \Lambda \) are increasing in \( \vartheta_I \). This captures the pro-competitive effect of creative destruction: by reducing equilibrium markups, creative destruction reduces misallocation and increases TFP and equilibrium factor prices holding the productivity frontier, \( Q_t \), fixed. Note also that the standard deviation of log markups is given by \( \theta^{-1} \) and is hence also decreasing in \( \vartheta_I \). As stressed in in the literature on misallocation, the dispersion in log TFPR co-moves with aggregate TFP.

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

\[
h_{\Delta+1}(a_P) = \frac{1}{\Delta!} (Ia_P)^\Delta e^{-Ia_P}.
\]

(18)

This implies that the average log markup in a market conditional on being served by the same firm for \( a_P \) years is given by

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

(19)

\[\] 10 De Blas and Russ (2015) analyze a version of Bernard et al. (2003), where the distribution of markups responds to trade policy. They consider a static model but allow for a finite number of firms, who differ in their efficiency (drawn from an exogenous Fréchet distribution). They show that the distribution of markups depends on the number of competing firms and that a higher number of competitors decrease markups.

\[\] 11 Note that \( \Delta \) is not a continuous variable but only takes integer values. For simplicity I treat markups as continuous when calculating \( M \) and \( \Lambda \) from \( G(\mu) \). See Section A-1.2 in the Appendix for the closed form expressions for the discrete case.
i.e. is increasing in age at a rate proportional to $I$. Hence, conditional on not being replaced, the distribution of markups continuously shifts outwards as incumbent firms engage in productivity improvements to rack up their monopoly power. This process of accumulation is faster, the higher $I$.

Markets, however, are not served by the same firm for eternity. In particular, creative destruction limits how long existing firms can survive. Because producers in a given market are replaced at rate $\tau$, the probability of serving a market for at least $a_P$ years is given by $e^{-\tau a_P}$. Hence, the extent to which firms can accumulate market power depends crucially on the degree of creative destruction $\tau$. If $\tau$ is high, it is rare to see firms serving a particular product market for a long time. The long-run distribution of markups is shaped by the interplay of these two processes, which lead to a pareto distribution.\(^{12}\)

### 2.6 The Firm-level Dynamics of Markups

Equations (18) and (19) characterize the life cycle dynamics of markups at the product level. These implications cannot be taken directly to the data because firms, i.e. the unit of observation where markups are commonly measured, are a collection of many products. This presence of multi-product firms makes the evolution of firm-level markups subtle.

Consider a firm of age $a_f$. On the one hand, old firms tend to have high markups for the reason encapsulated in (19): old firms are the only firms with the potential of having had enough time to build markups within a given market through a series of successful own-innovations. This “own-innovation channel” implies that markups and age should be positively correlated. On the other hand, old firms also had ample time to expand into new markets and lose products from their portfolio. And as markups in new, “marginal” products are lower than markups for the average variety the firm loses, this “creative destruction channel” tends to lower the extent to which markups increase in age. Hence, the model implies a product life cycle, where firms constantly accumulate market power in their existing products to increase profitability and add new, low markup products to their portfolio. This pattern is qualitatively consistent with the results reported in Foster et al. (2008).

To see this more clearly, suppose that firms were to never horizontally expand (i.e. $x = 0$) and hence never serve more than a single market. In that case, only the own-innovation channel is at play and the age of the firm $a_f$ directly corresponds to the time a market has been served by a particular producer, $a_P$. Hence, the average log markup by firm age is also given by (19). Allowing firms to expand horizontally into new markets breaks this tight link between markups and firm age. It is nevertheless the case that one can still derive an analytical characterization of the markup dynamics at the firm level.

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

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

where

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

\(^{12}\)To see this intuitively, suppose that (19) were to hold deterministically, i.e. $\ln \mu = \ln \lambda + \ln \lambda \times I a$. Then,

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

which is a pareto distribution. Jones and Kim (2016) exploit a similar structure to argue that creative destruction limits income equality by reducing the time entrepreneurs have to accumulate firm-specific human capital.
Notes: The figure displays the expected log markup as a function of age, i.e. $E[\ln \mu | a_f]$ given in (20).

**Figure 1:** The life cycle of Markups

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

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

Proposition 3 contains an analytic expression for the life cycle profile of markups. Note that, equation (20) has the same structure as (19), except that the mapping between firm age $a_f$ and product age $a_P$ is more complicated and depends on both the rate of incumbent expansion $x$ and the extent of creative destruction $\tau$. In particular, the possibility of firms breaking into new markets implies that $E[a_P|a_f] \leq a_f$. Moreover, it is easy to verify that $\lim_{x \to 0} E[a_P|a_f] = a_f$, so that (19) emerges as a special case.

The life cycle profile characterized in (20) is depicted in Figure 1. Surprisingly, the relationship is non-monotone. The intuition is the following. Recall that a given product is creatively destroyed with flow rate $\tau$. Hence, the average survival time for a given product is $\tau^{-1}$. For the set of very old firms, this is therefore exactly the average age of a given product in their portfolio, i.e. $\lim_{a_f \to \infty} E[a_P|a_f] = 1/\tau$. The limiting average markup for old firms (which is displayed in the red dashed line in Figure 1) therefore follows directly from (20) as $\ln \lambda (1 + I/\tau) = \ln \lambda (1 + \hat{\vartheta}_I^{-1})$.

Note that, as for the cross-sectional distribution, the churning intensity $\vartheta_I$ again emerges as a crucial determinant for the level of markups. If the churning intensity $\vartheta_I$ is high, the average markup of old firms is low.

The reason why the average markup for younger firms deviates from this level, is the result of selection. For young firms, the age of the products they sell is obviously negatively selected - a two year old producer cannot possible sell a product that has been around for four years. And because markups increase in the average age of firms’ product portfolios, young firms have low markups, which are expected to increase. Interestingly, once firms become sufficiently old, the expected age of the products they sell is positively selected. In particular, there is a chance that the firm still owns the product it initially started out with, which - for old firms - is older than the average product. In the limit, this effect vanishes as the probability that a 40 year old firm managed to cling on to its initial product for 40 years goes to zero.

Equation (20) also highlights how the life cycle dynamics of markups are shaped by firms’ endogenous innovation choices $(I, x)$ and the rate of creative destruction $\tau$. First of all, it is immediate that (for a given age) the average markup is increasing in $I$ as $E[a_P|a_f]$ is independent of $I$, but a higher rate of own-innovation allows firms to
increase their market power at a faster speed. The effects of incumbent expansion $x$ and creative destruction $\tau$ are more subtle. In the left panel of Figure 2 I depict the effect of an increase in the expansion rate $x$ (red line) and in the rate of creative destruction $\tau$ (blue line). Recall that $\tau = z + x$, i.e. holding $x$ fixed an increase in the rate of entry translates one-to-one into higher creative destruction. For visual clarity I focus on the early part of a firms’ life cycle, where markups are monotone in age. This is the empirically relevant case for my application.

While higher rates of incumbent expansion $x$ and creative destruction $\tau$ both decrease the extent of markup growth over the life cycle, the economics are very different. A higher rate of creative destruction reduces markup growth through higher churning: firms are unable to hold on to their products and hence do not have enough time to accumulate markups in the products they own. In contrast, a higher expansion rate $x$ reduces the average markup through a composition effect. If firms enter novel markets very frequently, only a small fraction of their sales is accounted for by old, high-markup products.

The patterns shown in Figure 2 suggest that - for a given rate of aggregate creative destruction $\tau$ - a higher expansion rate $x$ would reduce markups and hence increase allocative efficiency. However, as shown in Proposition 2, this is not the case as the distribution of markups is independent of $x$ conditional on $\tau$. To resolve this apparent contradiction, note that $x$ and $\tau$ also affect firms’ survival probabilities. Specifically, let $S(a_f)$ denote the share of firms surviving until age $a_f$. It can be shown that

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

(21)

where $\gamma(a)$ is given in Proposition 3. As seen in the right panel of Figure 2 (and shown formally in Section OA-1.2.3 in the Online Appendix), a higher rate of incumbent creative destruction $x$ increases the number of surviving firms at every age bin. Hence, while a faster rate of market expansion $x$ reduces the expected markup for a given age (left panel), it also increases the share of old firms, who have higher markups on average (right panel). In the cross-section, these two countervailing effects exactly cancel out, rendering the distribution of markups independent of $x$ conditional on $\tau$. In contrast, for the case of creative destruction, these two forces complement each other: not only is markup growth slower conditional on survival, but higher churning also implies that the age distribution

---

Figure 2: life cycle Dynamics of Markups and Survival

Notes: The figure displays the expected log markup as a function of age (see (20)) in the left panel and the cumulative survival probability (see (21)) in the right panel. The baseline model is shown in grey. The red (blue) line refers to an increase in incumbent expansion $x$ (creative destruction $\tau$).
shifts towards young, low-markup firms.

2.7 Sales and Employment Dynamics and the Distribution of Firm Size

In addition to the cross-sectional distribution of markups and the process of markup dynamics, the model also makes concise predictions for the dynamics of firm-size. As the stochastic process of firms losing markets and expanding into new markets is the same as in Klette and Kortum (2004), the firm size distribution (in terms of firm sales) takes exactly the same form. One major appeal of the Klette and Kortum (2004) framework is that it is consistent with many first-order features of the micro data. In particular the sales distribution is skewed, the variance of sales growth is decreasing in size, the probability of exit is declining in both size and age and Gibrat’s Law will be a good approximation for the growth for large firms (i.e. firms where the likelihood of exit is small). All these properties are true in my model too. Importantly, and in contrast to Klette and Kortum (2004), total sales and employment are no longer proportional. While firm sales are proportional to the number of active markets $n$, firm employment is also affected by the firm’s average markup.

Recall that the distribution of markups is fully characterized by a single endogenous statistic - the churning intensity $\varphi_I = \tau / I$. Similarly, the entire distribution of firm-sales is also fully characterized by a single endogenous statistic. Let $\varphi_x = x / \tau$ be the “expansion intensity”, i.e. the share of creative destruction, which is due to incumbent expansion. The equilibrium firm size distribution (in terms of sales) is fully determined by $\varphi_x$. In particular, the number of active firms $F$ is given by $F = \frac{1 - \varphi_x}{\varphi_x} \ln \left( \frac{1}{1 - \varphi_x} \right)$ and the share of aggregate sales accounted for by firms with at most $n$ markets, $\Omega_n$, is given by $\Omega_n = 1 - (\varphi_x)^n$. Hence, if incumbents’ expansion activities are an important component of the process of creative destruction, i.e. $\varphi_x$ is high, the number of active firms is small, average firm size is large and a large share of output is produced in large firms. Formally, both $F$ and $\Omega_n$ are decreasing in the expansion intensity $\varphi_x$.

Interestingly, the determinants of the markup and firm-size distribution neatly separate. The distribution of market power only depends on the speed of firm own-innovation relative to creative destruction, $\varphi_I$. In contrast, the distribution of sales is fully determined from the share of incumbent expansion relative to creative destruction, $\varphi_x$. This suggests that the dynamics of markups and firm size are exactly informative about the different margins of firm growth.

To see that this intuition is correct, let $E[\ln l_f | a_f]$ denote the expected log employment of firms of age $a_f$, i.e. the life cycle of firm size. As I show in Section A-1.4 in the Appendix, the theory implies that

$$E[\ln l_f | a_f] - E[\ln l_f | 0] = \frac{1 - \gamma(a)}{\gamma(a)} \sum_{j=1}^{\infty} \ln j \times \gamma(a)^j - \ln \lambda (I \times E[aP | a_f]),$$

where $\gamma(a)$ and $E[aP | a_f]$ are characterized in Proposition 3. The more important firms’ markup-increasing innovations $I$ relative to their expansion rate $x$, the steeper the age-profile of markups and the flatter the extent of life cycle employment growth. If there is no scope for incumbent own-innovation (as in Klette and Kortum (2004)), sales and employment are proportional at all ages as markups are constant. If in contrast firms would never expand

---

14 Another difference is that my model predicts a deviation from the exact proportionality between R&D spending and sales. In particular, it is easy to verify that the R&D intensity, i.e. R&D spending as a fraction of sales, of firm $f$ is proportional to $\varphi_x^{-1} x^{\zeta} + \varphi_x^{-1} I^{\zeta} = \frac{1}{\mu_f}$, where $\mu_f$ is the firm-level markup defined in (3). As I will discuss below, average markups are increasing in age (at least for the majority of the age distribution) and size. Hence, small, young, low-markup firms tend to spend relatively more on R&D. This is qualitatively consistent with the findings reported in Akcigit and Kerr (2018). In absolute terms, R&D spending is of course increasing in size, i.e. the model generates systematic differences in R&D spending across firms.

15 The mass of firms serving $n$ markets is given by $\omega(n) = \frac{1}{n} \left( 1 - \varphi_x \right) \varphi_x^{n-1}$. These results are derived in Section A-1.2 in the Appendix.
vertically, i.e. \( x = \partial_x = 0 \), firms only serve a single market irrespective of their age and sales would not be a function of age.\(^\text{16}\) As markups increase in age, firm employment would - conditional on survival - decrease in age while markups and profitability increase. Hence, the relative speed at which firms increase their markups relative to their size identifies the relative importance of firms expanding their scope of production horizontally, i.e. by creatively destroying novel products, or vertically, i.e. through productivity improvements in their existing markets.

### 2.8 Comparative Statics: The Effects of Entry Costs and Market Barriers

The results above highlight that markups, misallocation, growth and the firm size distribution are jointly determined in equilibrium. In particular, all these outcomes directly depend on the efficiencies of the innovation, expansion and entry technology \((\varphi_f, \varphi_x, \varphi_z)\). In this paper I apply the theory to the manufacturing sector in Indonesia. More specifically, I study the implications of a particular friction, which has been argued to be important in developing countries: firms might face high costs to enter new product markets. My theory highlights that such costs come in two flavors. They could either be market barriers, i.e. frictions for existing firms to expand into new product markets, or entry costs, i.e. costs for new firms to enter the economy.\(^\text{17}\) It turns out that they have qualitatively different implications.

For simplicity I model both of these frictions in a reduced-form way as being subsumed in the expansion and entry efficiency terms \( \varphi_x \) and \( \varphi_z \). Consider for example the case of incumbent creative destruction \( x \). The cost shifter \( \varphi_z \) parametrizes the combined resource requirements of coming up with the technology of producing a superior product and successfully introducing it in the market. Hence, a low level of \( \varphi_z \) can for example reflect license requirements or bureaucratic red tape, which have to be overcome before a firm can be active in a new market, or other policies aimed to shield existing producers from potential competitors.\(^\text{18}\) Similarly, a low level of \( \varphi_z \) can also be seen as a stand-in for lengthy approval processes for new producers to enter the market. The notion of high market barriers or high entry costs is therefore isomorphic to low expansion or entry efficiency, i.e. costs for new firms to enter new product markets, or entry costs, i.e. costs for new firms to enter the economy.\(^\text{19}\) The effects of these frictions on the extent of misallocation and the firm-size distribution is summarized in the following Proposition.

**Proposition 4.** Consider a stationary equilibrium and suppose that \( \zeta \geq \bar{\zeta} > 1 \). Higher entry costs reduce the churning intensity \( \vartheta_1 = \tau/I \) and increase the expansion intensity \( \vartheta_2 = x/\tau \). Higher market barriers reduce both the churning intensity \( \vartheta_1 \) and the expansion intensity \( \vartheta_2 \). Formally,

\[
\frac{\partial \vartheta_1}{\partial \varphi_z} > 0 \quad \frac{\partial \vartheta_2}{\partial \varphi_z} < 0 \quad \text{and} \quad \frac{\partial \vartheta_1}{\partial \varphi_x} > 0 \quad \frac{\partial \vartheta_2}{\partial \varphi_x} > 0.
\]

Higher market barriers and entry costs therefore increase misallocation, i.e. reduce \( M \) and \( \Lambda \). In contrast, market barriers (entry costs) reduce (increase) average firm size \( F^{-1} \) and the output share of large firms \( \Omega_n \). While both entry costs and market barriers reduce creative destruction, the effect on the aggregate growth rate \( g \) is ambiguous.

\(^{16}\)Note that \( \lim_{\varphi_x \to 0} F = 1 \), i.e. there is a unit continuum of firms producing one product each.

\(^{17}\)One widely used measure of entry costs is developed in Djankov et al. (2002). They measure the fees and time costs to legally operate a business for a variety of countries. Such variation in the regulation of entry has been linked to cross-country income differences in Barseghyan (2008), Barseghyan and DiCecio (2009) or Herrendorf and Teixeira (2011). There are also studies focusing on particular episodes of delicensing. The dismantling of India’s Licence Raj, for example, has been studied in Aghion et al. (2008). Even though all these studies refer to “entry costs”, the empirical variation is likely to capture both market and entry barriers in the sense of my theory.

\(^{18}\)Bento (2016) for example argues that product market regulation in the US is much more accommodating to competition than in many poor countries.

\(^{19}\)More formally, let \( \phi_x \leq 1 \) ( \( \phi_z \leq 1 \) be the probability that an expanding firm (entering firm) replaces the existing producer conditional on having generated a superior technology. Define the realized expansion and entry flows rates above as \( x = \phi_x \bar{z} \) and \( z = \phi_z \bar{x} \), where \( \bar{x} \) and \( \bar{z} \) are the gross expansion and entry rates. The model above incorporates these frictions once we define the expansion and entry cost shifter \( \varphi_x \) and \( \varphi_z \) as \( \varphi_x = \hat{\varphi}_x \phi_x \bar{z} \) and \( \varphi_z = \hat{\varphi}_z \phi_z \bar{x} \), where \( \hat{\varphi}_x \) and \( \hat{\varphi}_z \) denote the technological efficiency of the entry and expansion technology.
Proof. See Section A-1.5 in the Appendix. The restriction that $\zeta \geq \overline{\zeta}$ is a sufficient condition. It can be shown that $\overline{\zeta} < 2$. \hfill \square

Proposition 4 summarizes the consequences of market barriers and entry costs on the stationary equilibrium in this economy. Crucially, frictions for existing firms to enter new product markets and higher costs for new firms to enter the economy have qualitatively different implications. While both type of frictions reduce the intensity of churning $\vartheta_I$ and hence increase misallocation, they affect the firm size distribution differentially. Market barriers for existing firms reduce the expansion intensity $\vartheta_x$ and therefore reduce average firm size and increase the number of firms. In contrast, higher entry costs increase the expansion intensity $\vartheta_x$, which leads to bigger but less firms being active in equilibrium.\(^{20}\)

This suggests that in order to explain the joint behavior of misallocation and firm size across countries, variation in market barriers for existing firms are theoretically attractive in that they naturally imply a negative correlation between misallocation and average firm size. This is qualitatively consistent with the cross-country evidence as firms in poor countries are small and misallocation is argued to be rampant. Entry barriers in contrast face somewhat of an uphill battle, because their first-order effect on important moments like average firm size or the number of producers is counterfactual: high costs of entry will increase misallocation but also make firms larger.

It is also noteworthy to point out that Proposition 4 does not contain a result on the relationship between entry costs or market barriers and the endogenous growth rate $g$. The reason is that this relationship is ambiguous. If creative destruction and own-innovation are strong substitutes, it is possible that $I + \tau$, which - recall - determines the equilibrium growth rate, increases in response to higher entry of market barriers. In fact, this substitutability is not only a theoretical possibility but turns out to be quantitatively important in the calibrated economy.\(^{21}\)

2.9 Theoretical Extensions

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

2.9.1 Stochastic Step Size

Consider first the assumption of a constant step size. It turns out that this assumption is not only easy to dispense with but that all my results directly apply to a more general environment. In particular, suppose that conditional on successfully innovating upon an own product or entering a new product market, the firm improves upon the existing producer by $\tilde{k}$ steps (each of size $\lambda$), where $\tilde{k}$ is a random variable with $p_k = P[\tilde{k} = k]$ and $\sum_{k=1}^{\infty} p_k = 1$. The baseline model is the special case with $p_1 = 1$. In Section A-1.6 in the Appendix I show that the model with this extension is as tractable as the baseline model. In particular, the value function can still be solved explicitly, the innovation and entry policies $(I, x, z)$ are constant and the stationary equilibrium can be explicitly characterized.

\(^{20}\)The effect on $\vartheta_I$ is subtle. Holding $I$ constant, higher market barriers or entry costs lower creative destruction and hence increase misallocation. As for the response of $I$ there are two effects. On the hand, the decline in creative destruction raises firms’ incentives to increase their markups. On the other hand, higher costs might reduce the share of production workers and thereby aggregate demand. It is nevertheless possible to show that the effect on the churning intensity $\vartheta_I = \frac{I}{\bar{z}}$ is unambiguously negative.

\(^{21}\)While this result sounds similar to the findings in Aghion et al. (2001) and Aghion et al. (2005), who argue that product market competition increases growth through higher innovation incentives for incumbent firms, the mechanism is different. In my model, the ambiguous effect on the aggregate growth rate is a composition effect, whereby an increase in market barriers (entry costs) reduces firm expansion $x$ (entry $z$) but increases firms’ incentives to raise markups $I$.
Importantly, the link between the innovation environment and the endogenous distribution of markups is very similar as in the baseline model.

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

1. The unique stationary distribution of quality gaps \( \{\nu(\Delta)\}_{\Delta = 1}^{\infty} \) is defined recursively as
   \[
   \nu(\Delta) = \frac{1}{1+\vartheta_I} \left( \sum_{m=1}^{\Delta-1} \nu_m p_{\Delta-m} \right) + \frac{\vartheta_I}{1+\vartheta_I} p_j \quad \text{for } \Delta = 1, 2, 3...
   \]

   Hence, the churning intensity \( \vartheta_I = \tau/I \) is still a sufficient statistic for the distribution of quality gaps. In particular, a higher churning intensity \( \vartheta_I \) decreases the distribution of markups \( G(\mu; \vartheta_I) = \sum_{j=1}^{\ln \mu/\ln \lambda} \nu_{\Delta} \) in a first-order stochastic dominance sense, i.e.
   \[
   \vartheta_I^H > \vartheta_I^L \Rightarrow G(\mu; \vartheta_I^H) > G(\mu; \vartheta_I^L)
   \]

2. Suppose that \( p_n = \frac{1-\kappa}{\kappa^n} \), where \( \kappa < 1 \). Define \( \theta(\kappa) = \frac{1}{\ln \Lambda} \ln \left( \frac{1+\vartheta_I}{1+\kappa \vartheta_I} \right) \). As in Proposition 2, the distribution of markups, the degree of misallocation \( \mathcal{M} \) and the labor wedge \( \Lambda \) 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 Appendix.

Proposition 5 shows that the results from the baseline model apply in a straight-forward way. Importantly, the implications for the distribution of markups and the endogenous degree of misallocation are unchanged. It is still the case that the distribution of markups is fully determined from the innovation intensity \( \vartheta_I \) and that creative destruction is pro-competitive. The special case of \( p_n \propto \kappa^n \) is particularly instructive.\(^{22}\) For this specification, one can show that the endogenous distribution of markups is again Pareto with shape parameter \( \theta(\kappa) \). As expected, holding \( \vartheta_I \) constant, the shape parameter is decreasing in \( \kappa \), i.e. a more dispersed exogenous step-size distribution results in a more dispersed markup distribution in equilibrium. And as in the baseline model, an increase in \( \vartheta_I \) increases the pareto tail. Because the aggregate consequences of misallocation, \( \mathcal{M} \) and \( \Lambda \) only depend on the pareto tail of the markup distribution, the same formulas as in the baseline directly translate. Note that the baseline model is nested as the case of \( \kappa = 0 \).

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

\(^{22}\)If \( p_n \propto \kappa^n \), the requirement that \( \sum_{n=1}^{\infty} p_n = 1 \) implies that \( p_n = \frac{1-\kappa}{\kappa^n} \).
2.9.2 CES-Preferences

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

\[ Y_t = \left( \int y_{it}^{\sigma-1} di \right)^{\frac{1}{\sigma}}, \]

where \( y_{it} \) is the amount of variety \( i \), which as before can be produced by multiple firms, i.e. \( y_{it} = \sum_{j \in S_i} y_{f_{it}} \).

While I relegate a detailed analysis to Section OA-1.5 in the Appendix, I here highlight the main reasons, why the baseline case of \( \sigma = 1 \) simplifies the analysis and which results carry over to this more general case.

First of all, if consumers’ preferences were to take the more general CES form, equilibrium markups would not necessarily be determined through Bertrand competition as firms with a sufficiently large productivity advantage \( \Delta \) might prefer to charge the usual CES markup instead of the limit price. Formally, the equilibrium markup for variety \( i \) is given by

\[ \mu_i = \mu(\Delta_i) = \min \left\{ \frac{\sigma}{\sigma-1}, \lambda^\Delta \right\}. \quad (23) \]

Secondly, while it is still the case that output can be written as \( Y_t = Q_t M_t L_t P_t \), the misallocation term \( M_t \) and the labor wedge \( \Lambda_t \) now take the form

\[ M_t = \left( \frac{\int \mu(\Delta_i)^{-\sigma} di}{\int \mu(\Delta_i)^{-\sigma} di} \right)^{\frac{1}{\sigma-1}}, \quad \Lambda_t = \left( \frac{\int \mu(\Delta_i)^{-\sigma} di}{\int \mu(\Delta_i)^{-\sigma} di} \right)^{\frac{1}{\sigma-1}}, \quad (24) \]

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

\[ \pi(q_i, \Delta_i) = \left( 1 - \frac{1}{\mu(\Delta_i)} \right) \mu(\Delta_i)^{1-\sigma} q_i^{\sigma-1} w^{1-\sigma} Y_t. \quad (25) \]

Equations (23) - (25) highlight why the case of \( \sigma = 1 \) is particularly tractable. First of all, if \( \sigma > 1 \), the misallocation wedge \( M_t \) (and hence all aggregate outcomes) depend on the joint distribution of quality \( q \) and quality gaps \( \Delta \). If \( \sigma = 1 \), only the marginal distribution of quality gaps \( \Delta \) is required. Secondly, both the level of quality \( q \) and the quality gap \( \Delta \) determine the level of profits \( \pi_t \) and hence are state variables for the firm’s dynamic programming problem. This also implies that profits (and hence also sales and employment) depend explicitly on the level of quality \( q \) so that the endogenous distribution of \( q \) is required to calibrate the model to firm-level data. In particular, the distribution of \( q \) (appropriately scaled) has to be stationary for the implied distribution of firm-level sales to be stationary. This was not required in the baseline model where \( \sigma = 1 \).\(^{23}\)

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

\[ c_l(I; \Delta, q) = \left( \frac{q}{Q_l} \right)^{\sigma-1} \frac{1}{\varphi_l} I^\varsigma, \quad (26) \]

i.e. the cost of innovation scale at the same rate in \( q \) as firm profits. This requirement, which is for example also employed in Atkeson and Burstein (2010), is required to ensure that the model is consistent with Gibrat’s Law, i.e.

\(^{23}\)In fact, the quality distribution in the baseline model is not stationary, even though firm-level employment, sales and profits are.
that growth rates are independent of firm size, at least for large firms.\textsuperscript{24}

**Proposition 6.** Consider the model with CES demand and let the cost of own-innovation be given by (26). Along a BGP equilibrium:

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

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

$$\nu(\Delta) = \frac{\tau}{I(\Delta)} \prod_{m=1}^{\Delta} \frac{1}{1 + \frac{\tau}{I(m)}}.$$  \hfill (27)

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

$$\mathcal{M} = \frac{\left(\sum_{\Delta=1}^{\infty} \mu(\Delta)^{1-\sigma} \nu(\Delta)\right)^{\frac{\sigma}{\sigma - 1}}}{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{-\sigma} \nu(\Delta)}, \quad \text{and} \quad \Lambda = \frac{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{-\sigma} \nu(\Delta)}{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{1-\sigma} \nu(\Delta)}. \quad \hfill (28)$$

**Proof.** See Section OA-1.5 in the Appendix. There I explicitly characterize the value function along the BGP, show that firms’ own-innovation $I$ is independent of $q$ and derive (27).

The main result of Proposition 6 is that the endogenous distribution of quality gaps $\Delta$ 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 $\Delta$. If, however, the cost function takes the form in (26), one can show that the value function is homogeneous in $q^{\sigma - 1}$ and that the optimal own-innovation policy is independent of quality $q$ and consistent with Gibrat’s Law. Moreover, the rate of creative destruction $\tau$ is constant along a BGP. This implies that the distribution of quality gaps is determined by a set of differential equations akin to (16) in the baseline model, which has the solution (27). As before, only the ratios $\{\tau/I(j)\}_j$ are required to solve for the distribution of quality gaps and hence markups. If $I$ was constant, (27) is exactly the same solution as for the baseline model. And because markups are still fully determined from firms’ quality advantage (see (23)), the main economic insight from the baseline model is preserved in this more general environment: the distribution of markups is fully determined from the rate of creative destruction $\tau$ relative to firms own-innovation incentives $\{I(j)\}_j$.

The endogenous independence of productivity $q$ and quality gaps $\Delta$ is particularly attractive because it implies that the marginal distribution of productivity $q$ is not required to solve the model. In particular, the misallocation wedges along the BGP given in (28) follow directly from (24) and (27). These expression also highlight how the demand elasticity $\sigma$ affects the aggregate losses of misallocation $\mathcal{M}_t$. Holding $\mu(\Delta)$ and the distribution $\nu_\Delta$ fixed, the aggregate costs of heterogeneous markups tend to be increasing in $\sigma$ (see Hsieh and Klenow (2009)). However, two counteracting forces are also at play. First of all, a higher demand elasticity mechanically reduces markup dispersion and hence increase allocative efficiency - the higher $\sigma$, the more products will charge a markup of $\frac{z}{\sigma - 1}$ holding the distribution of $\Delta$ constant. Additionally, changes in $\sigma$ also affect firms’ innovation and entry incentives and hence the endogenous distribution $\nu_\Delta$. In Section 3.3 I show quantitatively that the results of the baseline model are not very sensitive to the choice of $\sigma$.

\textsuperscript{24}The cost function in (26) does not depend on the productivity gap $\Delta$. This is for simplicity. As long as the cost function takes the form $c(I, \Delta, q) = \left(\frac{\varphi \lambda}{\sigma z} \right)^{\frac{1}{\sigma - 1}} c(I, \Delta)$, with $c(I, \Delta)$ being convex in $I$, one can show that the optimal rate of own-innovation is independent of $q$. The specification in (26) has the benefit that the resulting innovation behavior of large firms is independent of both $q$ and $\Delta$, i.e. consistent with Gibrat’s Law.
3 Quantitative Analysis

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

3.1 Data

The main data set for the empirical analysis is the Manufacturing Survey of Large and Medium-Sized Firms in Indonesia (Statistik Industri). This data has also been used in Amiti and Konings (2007), Blalock et al. (2008), Yang (2012) and Hsieh and Olken (2014). The Statistik Industri is an annual census of all formal manufacturing firms in Indonesia and contains information on firms’ revenue, employment, capital stock, intermediate inputs and other firm characteristics. I will focus on the time period between 1990 and 1998, i.e. the years prior to the Indonesian financial crisis. My final sample has about 180,000 observations.

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

<table>
<thead>
<tr>
<th>Firm size distribution</th>
<th>Entrants</th>
<th>Exiting firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Quantiles 25% 50% 90%</td>
<td>Entry rate Share of employment sales</td>
<td>Exit rate Share of employment sales</td>
</tr>
<tr>
<td>143 27 45 351</td>
<td>10.4% 5.0% 3.9%</td>
<td>8.2% 4.4% 3.3%</td>
</tr>
</tbody>
</table>

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

Table 1: The Manufacturing Sector in Indonesia

For the purposes of this paper, this focus on large producers has advantages and disadvantages. On the positive side, this data covers firms for which considerations of productivity improvements and strategic pricing are more relevant. The majority of micro-firms in Indonesia are arguably subsistence entrepreneurs, which are unlikely to engage in such activities and which do not compete in the same product markets as large, formal employers.\(^{27}\)

\(^{25}\)To be absolutely precise, the data is collected at the plant level. As the majority of plants reports to be single branch entities, I will for the following refer to each plant as a firm.

\(^{26}\)See Table OA-2 in Section OA-2.1 in the Online Appendix for details.

\(^{27}\)There is mounting evidence for the importance of “stagnant” entrepreneurs in developing economies - see e.g. Schoar (2010), Hurst and Pugsley (2012), Akcigit et al. (2015) or Hsieh and Olken (2014). That these firms do not compete in the same product markets as...
Additionally, the data has a panel dimension, which allows me to use the information contained in the dynamics of markups and firm size over the life cycle to calibrate the structural parameters. Hence, I will not have to rely on the cross-sectional age-size or age-markup relationship to test the implications of the theory. To the best of my knowledge, there is no dataset covering the universe of Indonesian firms, which has a panel dimension.

The main drawback of this selection criterion is that it complicates the measurement of entry and exit as I only observe firms appearing and disappearing from the data. Table 1 shows that there are on average 10.4% firms entering and 8.2% of firms exiting the data. Naturally, these firms are much smaller than the average firm so that the population of entrants (exiting firms) accounts for 5% (4.5%) of aggregate employment in the data. Interestingly, they account for an even smaller fraction of sales in the economy, reflecting the fact that they have smaller markups (as predicted by theory).

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

### 3.2 The Markup Life Cycle of Indonesian Firms

The main theoretical advance of this paper is to construct a dynamic model of firm-dynamics with endogenous markups. The theory stresses in particular that markups should systematically vary with firm age. In this section I provide direct evidence that the variation of markups of Indonesian manufacturing firms is consistent with the theory.

#### Measuring Markups

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

Because I mainly focus on the life cycle properties of market power, I do not need to estimate the level of markups. This implies that I can rely on a measure of markups, which does not require an estimate of firms’ production functions (or more precisely the output elasticities). To see why, consider a firm \(f\), which is a price-taker in input markets. The optimality conditions from the firms’ cost-minimization problem imply that the markup satisfies the equation

\[
\mu_f = \alpha_{l,f} s_{l,f}^{-1},
\]  

(29)

where \(\alpha_{l,f} = \frac{\partial \ln y_f}{\partial \ln l}\) is the output elasticity of labor and \(s_{l,f} = \frac{w_l}{p_y}\) is the firm’s labor share in value added (or more generally, any expenditure share of a flexible input).\(^{28}\) In my model, the output elasticity of labor is unity, so that

their larger, formal counterparts is e.g. argued in La Porta and Shleifer (2009) or La Porta and Shleifer (2014).

\(^{28}\)Note that the allocative markup \(\mu_{l,f}\) depends on the payment share of production workers relative to sales. Empirically, I cannot precisely distinguish between production and innovation workers. The theory implies that the labor share of the entire workforce at firm \(f\), \(s_{l,f}^{total}\), is given by

\[
s_{l,f}^{total} = \frac{w}{\sum_{i \in N_f} \left[ l_f (\Delta_i, q_i) + l_t (\Delta_i, q_i) + l_x (\Delta_i, q_i) \right]} = \left( 1 + \frac{w}{p_y} \phi_f \phi_s \phi_x \right) \mu_f^{-1} + \frac{w}{p_y} \phi_s \phi_x. \]  

23
(29) implies (3). Note that the derivation of (29) did not use any information on the structure of demand or how firms compete.

If \( \alpha_l \) was known, one could directly infer firms’ markups from their observed labor shares. If \( \alpha_l \) is not known, but assumed to be constant across firms (i.e. stemming from a Cobb-Douglas production function), (29) 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 \( \mu_f \) is the residual from the regression

\[
\ln s^{-1}_{i,ft} = \delta_s + \delta_t + u_{ft},
\]

i.e. \( \ln \hat{\mu}_{ft} = \ln \hat{u}_{ft} \). Here, \( \delta_s \) is a set of 5-digit industry fixed effects and \( \delta_t \) is a set of year fixed effects. Under the assumption that \( \alpha_{lf} \) does not vary within 5-digit industries, the age variation of \( \hat{\mu}_{ft} \) is exactly the same as if I had estimated \( \alpha_l \) at the 5-digit level in a first stage and then calculated \( \mu_f \) according to (29) using the estimated \( \hat{\alpha}_l \). However, I also consider richer specifications, where I explicitly control for firms’ input choices like the capital- or material intensity, to allow for additional variation in output elasticities across firms within 5-digit industries. Furthermore, I also report results where I measure markups from firms’ material shares instead of labor shares.

Results

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

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

Because \( \frac{w}{Y} \frac{\partial f}{\partial z} x \) and \( \frac{w}{Y} x \) is constant, the variation in \( s^{Total}_{l,f} \) across firms and along the life cycle is entirely driven by the variation in \( \mu_{f}^{-1} \).

Furthermore, my data is standard in the sense that it does not contain information on firm-specific prices. Hence, to estimate the output elasticity \( \theta \) (which corresponds to physical output) one needs to impose additional structure.
Notes: The figure shows the life cycle of markups. To calculate the markup life cycle, I focus on the unbalanced panel of firms entering the economy after 1991. I calculate log markups within 5-digit-industry-year cells, then calculate the average by the age of the cohort and normalize log markups of entering cohorts to zero. Because of attrition, the size of the cohort is declining in age. The dots reflect the size of the cohort. I also depict the 90% confidence intervals around the estimated average profile.

Figure 3: The Life Cycle of Markups in Indonesia

To study the life cycle profile of markups more systematically, let me provide some additional evidence for the patterns in Figure 3. I will focus on regressions of the form

\[ \ln(\mu_{ft}) = \delta_t + \delta_s + \beta \times \text{age}_{ft} + \varrho \times \ln(k_{ft}/l_{ft}) + h_{ft} + \gamma + u_{ft}, \]  

(31)

where \( k/l \) denotes the firms’ capital-labor ratio, \( h \) contains additional firm-characteristics and \( \delta_t \) and \( \delta_s \) 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. Hence, the coefficient on age is identified from the rate of life cycle growth and not from the cross-sectional relationship between markups and age. The results are contained in Table 2.

<table>
<thead>
<tr>
<th>Dependent variable: log markup</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0137***</td>
<td>0.0103***</td>
<td>0.0106***</td>
<td>0.0134***</td>
<td>0.0108***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00122)</td>
<td>(0.00117)</td>
<td>(0.00116)</td>
<td>(0.00158)</td>
<td>(0.00127)</td>
<td></td>
</tr>
<tr>
<td>Entry</td>
<td>-0.0250***</td>
<td>0.00777</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00638)</td>
<td>(0.00744)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exit</td>
<td>-0.0176**</td>
<td>-0.0144*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00835)</td>
<td>(0.00836)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Control for capital-intensity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( N )</td>
<td>55212</td>
<td>55212</td>
<td>55212</td>
<td>48556</td>
<td>48556</td>
<td>42434</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.219</td>
<td>0.287</td>
<td>0.305</td>
<td>0.293</td>
<td>0.294</td>
<td>0.298</td>
</tr>
</tbody>
</table>

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

Table 2: The Life Cycle of Markups in Indonesia

The first column contains the specification displayed in Figure 3. On average, markups increase by roughly
1.4% per year. In columns 2 and 3, I include firms’ capital-labor ratio to control for a correlation between capital-intensity and firm size (and hence age). Doing so reduces the estimated age-coefficient slightly. In column 2 I simply include the log of firms’ capital-labor ratio, in column 3 I control in a less restrictive way by including 50 fixed effects for 50 quantiles of the distribution of capital-labor ratios. In column 4, I show that both entrants and exiting firms have lower markups. This is consistent with the model: both entering and exiting firms are small and therefore - on average - young.30 In column 5, I include firm age directly, which should account for the low markup of both entrants and exiting firms. For the case of entrants, this is indeed born out by the data. The results for exiting firms are not consistent with theory. In the model, exit is only a function of the number of markets \( n \) and not the age of the firm. However, the model implies that markups are decreasing in size holding age fixed. Intuitively: by having been unsuccessful in expanding into new markets, small old firms did not add low markup market into their portfolio. Hence, conditional on age, exiting firms should actually have higher markups. While the negative coefficient does increase once age is controlled for, exiting firms still have lower markups given their age. In column 6 I directly control for selection by conditioning on survival. In the theory, there is no selection in that the distribution of markups conditional on age is the same for all firms. In the data, the growth of markups shown in Figure 3 could stem from a higher exit hazard of firms, that systematically have low markups. Column 6 shows that markups are increasing over time even for those firms that do survive until the end of the sample.31

**Alternative misallocation frictions** My theory abstracts from any distortions to firms’ input choices. This allows me to use the data on firms’ factor shares to measure markups. This interpretation might be misleading if firms are subject to frictions, which distort their input choices. Because I am mostly interested in the change of markups over the firms’ life cycle, input distortions, which are constant at the firm-level do not invalidate the results in Table 2. More problematic are frictions, which systematically change in firm age or size. The fact that revenue productivity increases systematically in firm size forms the basis of a large literature on size-dependent policies (see e.g. Guner et al. (2008), Buera and Jaef (2016) or Bento and Restuccia (2017)). While firm-specific market power is a natural mechanism to generate this pattern, there could be other possibilities for why large firms appear to be constrained - see e.g. the discussion in Hsieh and Olken (2014).

In Section OA-2.6 in the Appendix, I present additional evidence on the comparison between factor shares reflecting markups or input distortions. In particular, I show that exporters and firms relying on FDI have in fact higher revenue productivity and that firms, who report to be capital-constrained, have lower revenue productivity. Hence, as often found in the literature, typical performance measures are positively correlated with marginal products. While this correlation is natural in environments where factor shares reflect market power, it is somewhat harder to rationalize with standard models of for example credit-constraints.32

---

30See also Foster et al. (2008), who - using US data - report qualitatively similar findings for particular industries selling homogenous products.

31In Section OA-2.6 in the Appendix I present additional robustness checks for these results. In particular, I show that the results do not substantially depend on whether or not I correct the measure of markups for measurement error as suggested in De Loecker and Warzynski (2012). I also consider the case of material shares in total sales. Markups should distort all factors within the firm equally, so that (29) should also hold for \( s_{M, ft} = \frac{m_{yt}}{p_{yt}} \), i.e. the share of sales going to materials. The results are very similar to the ones reported in Table 2 when firm’s material-labor ratio \( \ln (m/l) \) is controlled for. This is important, as there is a strong positive correlation between \( \ln (m/l) \) and age, i.e. older firms rely relatively more on materials. If this change in production factors is not controlled for, the correlation between material shares and age is negative.

32Benchmark dynamic models of financial constraints imply exactly the opposite pattern: borrowing constraints tend to bind in the early stages of the life cycle and get relaxed as the firms ages (see e.g. Clementi and Hopenhayn (2006)). However, in models of financial credit constraints, where firms face productivity shocks (see e.g. Buera et al. (2011), Moll (2014) or Midrigan and Xu (2014)), the relationship between factor shares and firm age is less clear. While old firms had more time to accumulate savings to overcome borrowing constraints, they might also have experienced a sequence of favorable productivity shocks, which increase their borrowing needs.
3.3 Markups, Misallocation and Growth in Indonesia

I now use a calibrated version of the model to quantify the macroeconomic consequences of market power.

Calibration

The model is very parsimonious. Given a rate of time preference $\rho$, which I set exogenously, the theory is fully parametrized by five parameters: the innovation step-size $\lambda$, the cost shifters for innovation, incumbent creative destruction and entry $\varphi_I, \varphi_z$ and $\varphi_x$ and the curvature of the innovation and expansion technology $\zeta$.

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

To identify the four structural parameters I use four moments. All of these moments have closed form expressions in the theory. First of all, I target the life cycle of markups, i.e. the average markup of 7 year old firm relative to entrants, given in (20). Secondly, I match the observed life cycle of employment, i.e. average employment of 7 year old firms relative to entrants, given in (22). Third, I target the entry rate, which is given by (see Section 2.7)

$$\text{Entry Rate} = \frac{z}{F} = \frac{z}{1 - \varphi_z \ln \left( \frac{1 - \varphi_x}{1 - \varphi_z} \right)} = \frac{x}{\ln \left( \frac{z + x}{z} \right)}.$$ 

Finally, I match a given rate of aggregate productivity growth, which is given by $g = \ln \lambda (\tau + I)$. These theoretical relationships allow me to write the four targeted moments directly in terms of the four unknowns $(I, \tau, z, \lambda)$. Calibrating the model therefore reduces to solving four non-linear equations. Also note the recursive structure of these equations. Given the empirically observed growth in markups, the entry rate and the employment life cycle are only dependent on $z$ and $x$. Given $z$ and $x$, $I$ and $\lambda$ are then uniquely determined from the aggregate growth rate and the markup life cycle.

Table 3 reports the results of this exercise. In terms of data moments, I require the model to match the fact that markups increase by 0.08 log points over a 7 year horizon (displayed in Figure 3) and that firms in Indonesia increase their employment by roughly 0.5 log points (i.e. a factor of 1.6) in the first 7 years of their life (see Figure 4 below, where I depict the employment life cycle from both the data and the model). Three to identify the flow rate of entry $z$, I require the model to be consistent with the observed entry rate of 10.4% reported in Table 1. Finally,

33Hence, in contrast to firms in India, which - according to Hsieh and Klenow (2014) - experience essentially no growth as they age, Indonesian firms do grow over time conditional on survival. However, both the sample of firms and the methodology underlying my calibration is different. First of all, the Indonesian data is biased towards bigger, formal firms. Secondly, my moments are estimated from panel data and not inferred from the cross-sectional age-size relationship. This turns out to be important as the cross-sectional age-size relationship in Indonesia is also relatively flat, which could be due to measurement error in firm age. See Section OA-2.6 in the Appendix, where I replicate the cross-sectional age-size relationship using the methodology by Hsieh and Klenow (2014). This difference between the “true” life cycle and the cross-sectional patters do not seem to be unique to the Indonesian context, but are also present in for example Chile (see Buera and Jaef (2016))
### Table 3: Calibration

I discipline $\lambda$ to match an aggregate rate of productivity growth of 3%.\(^{34}\) In Table 3 I report the data and model moments, the structural parameters and the implied endogenous innovation outcomes.

To further illustrate the mapping between the underlying structural parameters $(\varphi_I, \varphi_x, \varphi_z, \lambda)$, the equilibrium outcomes $(I, x, z, \tau)$ and the implied moments, Table 4 contains a sensitivity matrix and reports the change in equilibrium outcomes and moments for a 5% increase in the respective structural parameters. The results conform well with the economic intuition of my theory. First of all, an increase in the efficiency of innovation $(\varphi_I)$, expansion $(\varphi_x)$ and entry $(\varphi_z)$ raises own-innovation $I$, incumbent creative destruction $x$ and entry $z$ respectively. Furthermore, we see a sizable extent of crowd-out. Focusing first on entry and incumbent creative destruction, an increase in $\varphi_x$ reduces the equilibrium amount of entry $z$ and an increase in $\varphi_z$ lowers incumbent creative destruction $x$. Similarly, both of these parameters also negatively affect incumbents’ own innovation incentives, as the higher rate of creative destruction increases the effective discount rate of existing firms. That neither $\varphi_I$ nor $\lambda$ affect the equilibrium level of creative destruction by incumbents is apparent from (14). Note that a higher efficiency of own-innovation $\varphi_I$ increases the amount of entry by increasing the value of firms.

The implied changes in the resulting moments are reported in the lower panel of Table 4. These are consistent with the theoretical life cycle patterns shown in Figure 2. An increase in the efficiency of own-innovation $\varphi_I$ increases markup growth and reduces employment growth. The latter is the combination of both rising markups and lower sales growth as creative destruction $\tau$ rises and the expansion rate $x$ is unaffected. Increases in the efficiency of incumbent creative destruction $\varphi_x$ reduce the extent of life cycle markup growth as firms add new, low-markup products to their portfolio at a faster rate. In contrast, higher entry efficiency increases markup growth, despite lowering incumbent own innovation $I$ and increasing creative destruction $\tau$. The reason is that it also discourages creative destruction of incumbents. The effect of the step size $\lambda$ is also intuitive: the rate of growth, the entry rate and the extent of markup growth increase and the slope of the employment life cycle declines. Overall, as shown on the diagonal of the lower panel of Table 4, the respective parameters are strongly related to the moments one would expect them to be informative about: the innovation efficiencies $(\varphi_I, \varphi_x, \varphi_z)$ govern the life cycle of markups and employment and the entry rate and the step size parameter $\lambda$ affects the aggregate growth rate.

**Non-targeted moments** To get a better sense of the link between the model and data, I now also report a set of outcomes, which were not explicitly targeted. These are contained in Figure 4. In the top two panels I depict the calibrated and observed life cycle patterns for markups (left panel) and employment (right panel). While the

---

\(^{34}\) A detailed description of the construction of all these data moment is contained in Section A-2.2 of the Appendix. There I also present additional regression evidence for the life cycle of employment.
<table>
<thead>
<tr>
<th>Effects on endogenous outcomes</th>
<th>Change in ...</th>
<th>Initial level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of own innovation (I)</td>
<td>4.0% -1.1% -0.9% 1.9%</td>
<td>0.561</td>
</tr>
<tr>
<td>Incumbent creative destruction (x)</td>
<td>0.0% 5.0% -4.8% 0.0%</td>
<td>0.2078</td>
</tr>
<tr>
<td>Entry (z)</td>
<td>8.1% -21.9% 38.1% 24.3%</td>
<td>0.0326</td>
</tr>
<tr>
<td>Creative destruction (τ)</td>
<td>1.1% 1.4% 1.1% 3.3%</td>
<td>0.2404</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effects on equilibrium moments</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>life cycle of markups</td>
<td>3.6% -3.3% 0.6% 5.8%</td>
</tr>
<tr>
<td>life cycle of employment</td>
<td>-1.5% 6.9% -7.9% -3.5%</td>
</tr>
<tr>
<td>Entry rate</td>
<td>3.5% -7.1% 12.9% 10.2%</td>
</tr>
<tr>
<td>Growth rate</td>
<td>3.1% -0.3% -0.3% 7.3%</td>
</tr>
</tbody>
</table>

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

Table 4: Sensitivity Matrix

The model is only calibrated to match the data for 7 year old firms, it captures the general age pattern for both markups and employment reasonably well. In the lower panels of Figure 4 I focus on two additional predictions, namely the dynamic patterns of exit (left panel) and the cross-sectional size distribution (right panel). In equation (21) I derived the theoretical prediction for the probability of survival, $S(a)$. Through the lens of the model, which is stationary, I can measure $S(a_f)$ in two ways: either I can look at a cross-section at time $t$ and “count” the firms of different ages. Or I can follow a cohort through its life cycle and keep track of the surviving firms. In the lower left panel of Figure 4, I depict the model’s implication for $S(a)$ and the corresponding data - both measured from the cross-section in 2000 and from the evolution of the cohort, which entered in 1991. Two properties stand out. First of all, the survival probabilities estimated from the cross-section and from the panel turn out to be quite similar. Secondly, the model captures the dynamics of exit relatively well, even though it slightly over-predicts the extent of “shake-out”, i.e. firms exit at too fast a rate. The bottom right panel reports the fraction of aggregate output, which is accounted for by the smallest $x\%$ of firms, i.e. the Lorenz curve of the firm size distribution. Both in the model and the data the curve is below the 45-degree line, reflecting the dispersion in firm size. The empirical distribution is, however, more unequal than its theoretical counterpart, i.e. the model under-predicts the size advantage of large firms.\(^\text{35}\)

To further highlight the mapping between the theoretical mechanisms and the moments, in Section OA-2.3 in the Online Appendix I replicate Figure 4 for the calibrated model, where I abstract from the possibility of own-innovation. In this case, the model collapses to the baseline model of Klette and Kortum (2004). While this model implies by construction that markups are constant both across firms and across the life cycle, it can nevertheless successfully replicate the remaining three panels reported in Figure 4. This shows that it is exactly the information contained in the life cycle of markups which is important to discipline the extent of own-innovation.

The theory also makes predictions about the correlation of markups and size. While the model predicts that markups are increasing in firm size, the relationship is quantitatively small compared to the data. In Section A-2.3 in the Appendix, I derive an analytic expression for the average markup as a function of size, i.e. the size-based analogue to Proposition 3. For the calibrated parameters, the implied elasticity between markups and size is

---

\(^{35}\)In the model, firms’ only margin of employment growth is to enter in new markets and to replace other producers - productivity growth in existing markets actually reduce employment through increasing markups. This sharp distinction is conceptually and analytically useful. It is, however, restrictive. For example, if the elasticity of demand exceeded the Cobb-Douglas case of unity, increases in quality could also led to increases in employment as firms would pass-through some of their cost-reduction into prices consumers face. In that case, the model could rationalize a given slope of the age-employment schedule with margins other than creative destruction. See also Garcia-Macia et al. (2016) and Luttmer (2010).
Notes: The figure displays the model’s prediction and the data for the life cycle of markups (top left panel), employment growth (top right panel), survival (bottom left panel) and the concentration of value added (bottom right panel). Firm survival is measured both by the share of firms by age in the 2000 cross-section (relative to the number of firms at the time of entry) and the share of firms of the entering cohort in 1991 by age. For the employment life cycle see Section A-2.2 in the Appendix. To measure the distribution of firm size, I drop the highest and lowest 3 percent of firms to account for the potential of measurement error.

Figure 4: The Life Cycle, Survival and Output Concentration: Model vs Data
Markups and Misallocation

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$E[\mu]$</th>
<th>$\sigma(\ln \mu)$</th>
<th>$\sigma(\ln \mu_f)$</th>
<th>$M$</th>
<th>$\Lambda$</th>
<th>$I$</th>
<th>$x$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.5</td>
<td>11.8%</td>
<td>0.103</td>
<td>0.079</td>
<td>0.995</td>
<td>0.9</td>
<td>70%</td>
<td>26%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Notes: The table reports the endogenous tail parameter of the markup distribution $\theta$, 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 $M$ and $\Lambda$ (see (4) and (5)). It also shows the share of aggregate growth accounted for by $I, x$ and $z$, i.e. $\ln(\lambda)I/y$, $\ln(\lambda)x/y$ and $\ln(\lambda)z/y$. All results are based on the calibration reported in Table 3.

Table 5: Markups, Misallocation and Growth in Indonesia

positive, but very small: 0.006. Empirically, the elasticity is also positive but much larger: 0.23.

The reason why the model under-predicts the extent to which markups rise with size is the following. The only reason why larger firms should have higher markups, is that they are on average older. However, holding age fixed, the correlation between markups and size is negative. Consider, for example, firms that are too large given their age. These are firms, which got lucky in their expansion activities. But their markups are not systematically higher - in fact they are lower as successful expansion activities draw new, low markup products into the firm. In the calibrated model, this systematic negative correlation between size and market power given age 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. The correlation between size and markups would, for example, be stronger if some firms where more efficient to expand into new markets and would also have systematically higher quality draws, which would allow them to charge higher markups.

Implications for Markups, Misallocation and the Sources of Growth

I now turn to the aggregate implications of the calibrated model. In Table 5 I report the extent of misallocation and the drivers of growth in the stationary equilibrium. The equilibrium churning intensity $\vartheta_I = \tau_I$ implied by Table 3 is equal to 0.43. Proposition 2 therefore implies that the distribution of markups across products is pareto with shape $\theta = 9.5$. Average markups are therefore 12%. This in the range of the estimates of De Loecker and Warzynski (2012), albeit at the lower end. Depending on the specification used, they estimate markups between 10% and 28%. They are also lower than the results reported in De Loecker and Eeckhout (2017), who report a sales-weighted average markup of 20% in 1980 and 60% in 2012 for publicly traded firms in the US. In my model, average-markups are directly implied from the process of firm-dynamics and a given aggregate growth rate. Intuitively: for average markups to be higher (given the estimated life cycle slope) the step size $\lambda$ would need to be higher. This however, would also imply a higher aggregate growth rate.\(^{36}\)

In terms of markup dispersion, the calibrated model implies a standard deviation of log markups of about 0.1. While this dispersion is the welfare-relevant measure of markup heterogeneity, it does not directly compare to empirical measures of TFPR dispersion across firms. If firms are active in multiple product markets, the dispersion across firms is lower than the overall heterogeneity faced by consumers. This discrepancy is seen in Figure 5, where I depict the equilibrium distribution of markup across products (dark bars) and firms (light bars). It is clearly seen that the markup distribution across firms is compressed because it neglects the dispersion of markups within firms. Quantitatively, the dispersion across firms underestimates the actual dispersion by about 20%.

The static macroeconomic consequences of firms’ market power are summarized by $M$ and $\Lambda$. The model implies that TFP is lowered by 0.5% and wages are depressed by 10% relative to their social marginal product. The implied reduction in TFP seem small, especially compared to the much bigger numbers reported in Hsieh

\(^{36}\)Note also that in my model, markups are generated from a single mechanism: limit pricing. It hence abstracts from other reasons why markup could differ across firms like non-CES demand (see e.g. Edmond et al. (2018) or Dhingra and Morrow (2012)).
Notes: The figure shows the distribution of markup at the product level (dark bins) and at the firm level (light bins). The results are based on the calibration reported in Table 3.

Figure 5: The Stationary Distribution of Markups Across Products and Firms

and Klenow (2009). There are two reasons. First of all, the model only captures a small share of the observed dispersion in average revenue products across firms. Empirically, the standard deviation of \( \hat{u}_{ft} \) is 0.74, which is consistent with Hsieh and Klenow (2009) who find numbers of around 0.7 in China and India. The remainder could hence be explained by other frictions (e.g. adjustment costs as stressed in Asker et al. (2014)), model misspecification or measurement error. Secondly, Hsieh and Klenow (2009) consider an elasticity of substitution of three, whereas I impose a unitary demand elasticity. Recall that the change in aggregate TFP is approximately given by \( d \ln TFP = -\sigma^2 \text{dvar}(\ln TFPR) \), where \( \sigma \) is the elasticity of substitution across varieties. For \( \sigma = 1 \) and \( \text{dvar}(\ln TFPR) = 0.103^2 \), one exactly recovers a TFP loss relative to the efficient allocation of 0.5%.\(^{37}\)

The model also has implications for the origins of productivity growth and reallocation. Given the calibrated endogenous outcomes \( \tau \) and \( x \), the model implies that entrants account for \( \frac{13.5}{5} \approx 13.5\% \) of aggregate creative destruction. While on the low end, it is not entirely at odds with other findings in the literature. Foster et al. (2001, p. 309) for example find that about “15% of job creation is accounted for by entry and exit”. In terms of productivity growth, the calibrated model implies that the share of aggregate growth accounted for by firms’ own-innovation is given by about 70%. This is similar to Garcia-Macia et al. (2016), who estimate that about 75%-80% of incumbent growth is due to own-quality improvements. Note however, that own-quality improvements in Garcia-Macia et al. (2016) increase employment (as markups are assumed to be constant), while such productivity increases in my model are fully reflected in firms’ markups. These two facts imply that the share of aggregate growth accounted for by entering firms, \( \frac{1}{1+\tau} \), is low - it is only 4%. There are two main reasons why this is the case. First of all, I abstracted from the entry of new varieties. If new varieties are more likely to be produced by entering firms, the entry share in aggregate growth could increase. Secondly, I restrict the step size of innovation, \( \lambda \), to be the same across all sources. If entrants were to enter with technologies, which represented a drastic innovation, a given entry rate could be consistent with a larger share of growth.

\(^{37}\)The formula \( d \ln TFP = -\frac{\sigma^2}{2} \text{dvar}(\ln TFPR) \), used in Hsieh and Klenow (2009), relies on the assumption of physical productivity and \( TFPR \) be to independent and log-normally distributed. They estimate that \( \text{var}(\ln TPFR^{ND}) - \text{var}(\ln TPFR^{US}) \approx 0.24 \) For \( \sigma = 3 \), this implies a loss in aggregate TFP of 36%.
Stochastic Step Size and CES Preferences: Quantitative Implications

In Section 2.9 I showed how to generalize the theory along two dimensions: I allowed for the step size of successful innovations to be stochastic and I characterized the economy with CES preferences. In this section I show that these extension do not significantly change the quantitative results reported above.

Stochastic Step Size

To see that the restriction of a common step size is not restrictive, consider the model where the step size is drawn from \( p_n = \frac{1-\kappa}{\kappa} \kappa^n \) (see Proposition 5). The left panel of Figure 6 shows the extent of misallocation \( M \) as a function of \( \kappa \). The solid line correspond to the case where - for every \( \kappa \) - I recalibrate all parameters to match the exact same moments as in my baseline calibration. It is clearly seen that the line is essentially flat. Hence, as far as the aggregate implications for the degree of misallocation are concerned, heterogeneity in the number of quality steps is not particularly important.\(^{38}\) This is also seen in the right panel of Figure 6, which contains the endogenous distribution of markups and shows that the distribution for the stochastic step size model is very close to the baseline calibration.

Recalibrating the model is crucial to reach this conclusion. As seen in Proposition 5, the parameter \( \kappa \) obviously has a mechanical effect on the distribution of markups holding the churning intensity \( \vartheta_I \) fixed. In the left panel I depict the aggregate implications when I fix \( \vartheta_I \) at its baseline value: misallocation increases substantially and aggregate TFP is reduced by more than 10% for \( \kappa = 0.8 \). The reason is, of course, that the distribution of markups becomes much more dispersed when firms’ quality steps are stochastic but \( \vartheta_I \) is takes an parametric. In particular, as seen in the right panel, the share of products with markups exceeding 50% increases markedly.

The CES Model

The restriction to a unitary elasticity of substitution is potentially more consequential. It turns out, however, that the implications of the baseline model for the costs of misallocation are also quantitatively robust to changes in the demand elasticity.\(^{39}\) Recall that Proposition 6 established that the misallocation wedge

\(^{38}\)In Figure 6 I focus on \( M \) for expositional brevity. The same result holds for the labor wedge \( \Lambda \).

\(^{39}\)Section OA-1.5 in the Appendix contains all detailed derivations for the CES model.
along the BGP is given by

\[ M(\sigma) = \frac{\left(\sum_{\Delta=1}^{\infty} \mu(\Delta)^{1-\sigma} \nu_\Delta\right)^{\frac{1}{\sigma-1}}}{\sum_{\Delta=1}^{\infty} \mu(\Delta)^{1-\sigma} \nu_\Delta} \quad \text{where} \quad \mu(\Delta) = \min\left\{ \frac{\sigma}{\sigma-1}, \lambda^\Delta \right\}, \quad (32) \]

where the notation makes the dependence on \( \sigma \) explicit. The expression in equation (32) highlights that the demand elasticity \( \sigma \) has three effects. First of all, holding the distribution of quality gaps \( \nu_\Delta \) and the corresponding markups \( \mu(\Delta) \) constant, the demand elasticity \( \sigma \) determines how this heterogeneity is correctly aggregated. Secondly, \( \sigma \) directly determines the mapping from quality gaps \( \Delta \) to the markups firms actually post. The higher \( \sigma \), the lower the optimal CES-type markup \( \frac{\sigma}{\sigma-1} \). Hence, for a given distribution of quality gaps, a higher demand elasticity reduces both markup dispersion and the level of markups by truncating the right tail of the markup distribution. Finally, the distribution of quality gaps \( \nu_\Delta \) itself is endogenous and will change as a function of \( \sigma \).

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

\[ M^N(\sigma) = \frac{\left(\int \mu^{1-\sigma} dG(\mu)\right)^{\frac{1}{\sigma-1}}}{\int \mu^{-\sigma} dG(\mu)} = \frac{\theta + \sigma}{\theta} \left(\frac{\theta}{\theta + \sigma - 1}\right)^{\frac{1}{\sigma-1}} \quad (33) \]

I refer to this measures as the “naive” measure, because it abstracts from any feedback of the demand elasticity to the markups firms actually post. This expression is nevertheless helpful because it turns out that \( M^N \) is a relatively tight bound for the full effects captured by \( M(\sigma) \). In the left panel of Figure 7 I depict this naive measure in red. These naive losses from misallocation are increasing in \( \sigma \). Quantitatively, this mechanical effect can potentially increase the TFP cost of misallocation by about 1 percentage point. The actual misallocation losses \( M(\sigma) \), which are depicted in the dark lines in Figure 7 are smaller then their native counterparts. In particular, the model with a non-unitary demand elasticity can reduce TFP by about a percentage point, i.e. it “adds” about half a percentage point of TFP losses relative to the baseline model.\(^{40}\)

That the misallocation losses implied by the naive measure are an upper bound for the actual losses is intuitive. By truncating equilibrium markups at \( \frac{\sigma}{\sigma-1} \), a higher demand elasticity tends to lower misallocation by reducing the dispersion of markups. Additionally, it turns out that in the calibrated economy with elastic demand, the endogenous distribution of quality gaps \( \nu_\Delta \) is such that there is more mass on low markup products. This is for example seen in the right panel of Figure 7, where I depict the stationary distribution of markups for the baseline model and for the case of \( \sigma = 4 \). In the latter case, the highest markup in the economy is 33% so that there is no mass on the set products in the last two bins of the histogram. Importantly, the distribution of markups is also shifted to the left in that it puts more mass on products with markups lower than 10%. Hence, both the level of markups and their dispersion is lower. That the costs of misallocation are nevertheless slightly higher than in the baseline model is due to the CES aggregator.

These results for the stochastic step size model and the model with elastic demand suggest that the quantitative results stemming from the baseline model provide robust estimates for the extent of markup misallocation. More

\[ \Lambda(\sigma) = \sum_{\Delta=1}^{\infty} \mu(\Delta)^{-\sigma} \nu_\Delta \quad \text{and} \quad \Lambda^N(\sigma) = \frac{\int \mu^{-\sigma} dG(\mu)}{\int \mu^{1-\sigma} dG(\mu)} = \frac{\theta + \sigma - 1}{\theta + \sigma}. \]

As in Figure 7 it is again the case that the results of the baseline model are quantitatively robust to changes in \( \sigma \).

\(^{40}\)For brevity, in Figure 7 I focus on the misallocation wedge \( M(\sigma) \). Similar results hold for the labor wedge \( \Lambda(\sigma) \). The actual labor wedge \( \Lambda(\sigma) \) and the naive counterpart when the equilibrium outcomes are held fixed, \( \Lambda^N(\sigma) \), are given
Notes: The left panel shows the misallocation wedges $M$ and $M^N$ for different values of the demand elasticity $\sigma$. In the dark line I show the actual expressions (i.e. (32)) when the model is calibrated to the same moments as the baseline model. In the red line I depict the “naive” measure, which keep the distribution of markups at their baseline level (see (33)). The right panel shows the distribution of markups across products for both the baseline economy and for the economy with $\sigma = 4$.

Figure 7: Misallocation in the Model with CES Preferences

specifically: once the structural parameters are calibrated to match salient features of the extent of markup and employment growth and the aggregate entry rate, additional ingredients like step size heterogeneity or elastic demand do not seem to markedly affect the aggregate implications of markup heterogeneity.

3.4 Counterfactual Analysis: The Importance of Market Barriers and Entry Costs

How important are frictions like market barriers for existing firms or entry costs for new producers for the aggregate losses of misallocation, the economy-wide rate of productivity growth and the endogenous distribution of firm size?

In Proposition 4 I characterized these effects qualitatively. I now use the calibrated model and data from the US to provide a quantitative answer.

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

To implement this calculation I require information on the entry rate and the rate of life cycle growth in the US. As a benchmark for the rate of life cycle growth in the US, I build on the results of Hsieh and Klenow (2014), who report that US firms grow by a factor 2 at the 10 year horizon.\footnote{More specifically, they show that 10-14 year old plants are twice as large as plants less than 5 years old.} This corresponds to an annual rate of employment growth of about 7% conditional on survival. As for the entry rate, I target a value of 8% for the baseline results, which is consistent with Karahan et al. (2015), who report a start-up rate of between 8% and 11% for the whole economy and Akcigit et al. (2015), who calculate an entry rate of 7.5% in the US manufacturing sector. However, I will show explicitly how sensitive the results are with respect to these choices. For ease of comparison, I used the calibration to the Indonesia economy reported in Table 3 to also calculate the implied employment growth for firms...
### Calibration

<table>
<thead>
<tr>
<th></th>
<th>Indonesia</th>
<th>US</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry Rate</td>
<td>$z/F$</td>
<td>10.4%</td>
<td>8%</td>
</tr>
<tr>
<td>Employment life cycle</td>
<td></td>
<td>1.7</td>
<td>2</td>
</tr>
<tr>
<td>Expansion Barrier</td>
<td>$(\varphi_x / \varphi_x^{IND})^{-1}$</td>
<td>1</td>
<td>0.67</td>
</tr>
<tr>
<td>Entry Costs</td>
<td>$(\varphi_z / \varphi_z^{IND})^{-1}$</td>
<td>1</td>
<td>0.86</td>
</tr>
</tbody>
</table>

### Equilibrium Implications

#### Panel A: The Distribution of Firm Size

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of firms</td>
<td>$F$</td>
<td>0.313</td>
<td>0.122</td>
</tr>
<tr>
<td>Output share of small firms</td>
<td>$\Omega_1$</td>
<td>0.136</td>
<td>0.035</td>
</tr>
</tbody>
</table>

#### Panel B: Markups and Misallocation

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Life cycle of markups</td>
<td>$E[\mu]$</td>
<td>8.2%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Average markup</td>
<td>$\sigma(\ln \mu)$</td>
<td>11.73%</td>
<td>9.25%</td>
</tr>
<tr>
<td>Dispersion in log markups across products</td>
<td>$\sigma(\ln \mu_f)$</td>
<td>10.5%</td>
<td>8.47%</td>
</tr>
<tr>
<td>Dispersion in log markups across firms</td>
<td></td>
<td>8.1%</td>
<td>5.86%</td>
</tr>
<tr>
<td>Aggregate misallocation</td>
<td>$\mathcal{M}$</td>
<td>0.995</td>
<td>0.997</td>
</tr>
<tr>
<td>labor wedge</td>
<td>$\Lambda$</td>
<td>0.887</td>
<td>0.903</td>
</tr>
</tbody>
</table>

#### Panel C: Innovation, Expansion and Entry and Growth

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of growth</td>
<td>$g$</td>
<td>3%</td>
<td>2.91%</td>
</tr>
<tr>
<td>Own innovation</td>
<td>$I$</td>
<td>0.561</td>
<td>0.498</td>
</tr>
<tr>
<td>Incumbent creative destruction</td>
<td>$x$</td>
<td>0.207</td>
<td>0.267</td>
</tr>
<tr>
<td>Aggregate creative destruction</td>
<td>$\tau$</td>
<td>0.240</td>
<td>0.277</td>
</tr>
</tbody>
</table>

Notes: The first panel contains the calibration moments. The entry rate is simply the share of firms, which are entrants. For the US, the employment life cycle is calculated as average employment of firms between 10 and 14 years old relative to firms with age less than 5 and stems from Hsieh and Klenow (2014). The parameters for the Indonesian economy are contained in 3. For the US economy, I recalibrate the relative efficiency of expansion and entry, i.e. $\varphi_z$ and $\varphi_x$. All remaining parameters are the same as in Table 3.

Table 6: A Counterfactual: Changes in Market Barriers and Entry Costs

in Indonesia at the 10 year horizon, which turns out to be 1.7. Hence, the US entry rate is lower and the rate of life cycle growth higher.

The results of this exercise are contained in Table 6. In the top panel, I report the two new moments and the resulting estimates for the entry costs $(\varphi_z^{-1})$ and the market barriers $(\varphi_x^{-1})$. I express these estimates relative to the calibrated values for the Indonesian economy reported in Table 3. While both entry costs and market barriers are lower in the US, the market barrier margin is particularly important. The entry technology in the US is about 15% more productive, the costs for existing firms to break into new markets are about a third lower. The intuition is simple. For the model to generate a faster rate of life cycle employment growth, firms need to be willing to expand more aggressively. This calls for lower market barriers in the US. This however, reduces the equilibrium entry rate. In a stationary equilibrium, the mass of entrants has to be equal to the mass of exiting firms, i.e. the mass of single-product firms experiencing a creative destruction shock. Even though lower market barriers increase the rate of creative destruction, there is still less exit, simply because in an economy populated by large firms it is less likely that a producer who gets replaced in a particular product coincides with an exiting firm. To match the observed entry rate of 8%, more entry is required. Entry barriers in the US therefore also have to be lower.
In the remaining panels of Table 6 I report the equilibrium implications of this reduction in entry and expansion costs. I focus on three sets of results, which concern the distribution of firm size (Panel A), the distribution of markups (Panel B) and the sources of aggregate growth (Panel C). Panel A shows that the estimated reductions in entry costs and market barriers affect the firm size distributions markedly. As firms have more opportunities to expand their scale of production, the firm size distribution shifts to the right and the economy sustains less firms in equilibrium. The equilibrium number of firms declines by 60%, i.e. average firm size more than doubles. This reallocation comes especially at the expense of small firms so that the sales share of firms that serve only a single market, declines by more than 70%. Hence, seemingly small differences in the rate of entry and life cycle growth have very large effects on the cross-sectional distribution of firm size.

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

Finally, Panel C consider the implications on aggregate growth. The most striking result is that the economy-wide growth rate \( g \) hardly changes - if anything it slightly declines once entry barriers and expansion barriers are dismantled. The reason is the equilibrium effect on firms’ markup-increasing productivity investments. While creative destruction increases by 15%, firms’ incentives to increase productivity within their existing markets decline by 11%. That these different margins of growth are negatively related is not surprising (recall the optimality condition in (15), which showed that the marginal value to accumulate markups is discounted at rate \( \rho + \tau \)). This competition effect, which is present in most models of Schumpeterian growth, is sufficiently strong that \( I + \tau \) declines, even though \( \tau \) increases. In terms of aggregate GDP, holding technologies \( Q_t \) constant, the counterfactual decline in entry costs and market barriers raises GDP by 1.5%. This effect is the combination of lower misallocation and a reallocation of labor into the production sector.

The results in Table 6 have four important implications. First of all, seemingly large changes in the stationary firm size distribution and the number of active firms are fully consistent with empirically plausible small differences in observable entry rates, employment life cycle growth and the increase in markups by age. Secondly, such large differences do not imply that countries are predicted to grow at vastly different rates. A growth differences of the one reported in Table 6 only accumulates to a productivity level difference of 2% after 20 years. Hence, the model offers one explanation why firm size distributions across countries could be vastly different, while the distribution of income across countries might - for all practical purposes - be relatively stable. Third, the results suggest that frictions for existing firms to expand into new markets are more important than differences in entry costs to understand the empirical firm-level patterns across countries. While such frictions readily imply that many firms are small and experience little growth as they age, high entry costs would have the exact opposite implication. Finally, while misallocation is indeed predicted to be lower, the quantitative magnitude is relatively small, as the static

\[ \text{This is qualitatively consistent with the results reported in Hsieh and Klenow (2014), who argue that the increase in revenue-productivity by age, which in my model is proportional to markups, is steeper in India relative to the US. See also Bento and Restuccia (2017) or Fattal Jaef (2018).} \]
efficiency losses of markups are limited to begin with. To the extent that misallocation is an important determinant of cross-country differences in aggregate TFP, it is unlikely that differences in markup heterogeneity account for a substantial share of it.

Robustness  In Section OA-2.6 in the Online Appendix I examine the robustness of these results. First of all, I show that neither the implied changes in the firm size distribution, nor the implications for the aggregate growth rate depend substantially on the choice of the curvature parameter \( \zeta \). The resulting changes in the number of firms or the importance of small firms are all very similar to the ones reported in Table 6 and the quantitative effect of the changes in entry and expansion barriers on the equilibrium growth rate is very small (even though for large values of \( \zeta \), the predicted growth rate in the US is actually slightly higher). I also study the sensitivity with respect to the underlying moments, i.e. the extent of life cycle growth and the extent of entry. The baseline calibration for the US assumed an entry rate of 8% and that employment grows by a factor of two during the first 10 years of a firm’s life cycle. The elasticity of average firm size and and the share of small firms with respect to these two moments is quite sizable. If one were to for example assume that the extent of life cycle growth in the US was 2.5 instead of 2, the number of firms and the share of small firms would fall by 90%. In contrast, the effect on the growth rate is still almost indistinguishable from zero. In Section OA-2.8 in the Online Appendix I also report an alternative calibration, which does not rely on the observed entry rate but explicitly exploits the size cutoff of census data.

3.5 Market barriers and Entry Costs in Indonesia: Empirical Evidence

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

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

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

\[
y_{rst} = \delta_s + \delta_t + \beta \times \text{AvgSize}_{rst} + \gamma \times \ln (\text{pop}_r) + \alpha \times A_{gr} + u_{rst},
\]

where \( \delta_s \) and \( \delta_t \) are sector and time fixed effects, \( \text{AvgSize}_{rst} \) is the average size of producers active in region \( r \), in
Table 7: Firm size, Entry and Markups across Product Markets in Indonesia

| Sector s in time t and y_{r,t} are the different outcome variables mentioned above. Moreover, I also control for the size of the population in region r and the regional agricultural share to account for the effects of market size. Given the focus on the regional variation, I cluster all standard errors at the province level, to allow for correlation in the error term across industries within a province. 

My preferred measure of size is firms sales (rather than employment), as the theory predicts that average sales only depend on the expansion intensity $\vartheta_x = x^\beta$. To calculate the employment life cycle, I again rely on the panel dimension and adopt the same methodology as for Figure 4. To not identify the parameters from sparsely populated region-industry-year cells, I estimate (34) only for the set of observations, which contain at least 50 observations and I weight the regression by the number of observations in each bin.

Table 7 contains the results. In the first three columns I show that average firm size in a region is negatively correlated with entry and exit rates and positively correlated with the extent of life cycle growth. These correlations hold regardless of whether the source of variation across regions stems from entry costs or market barriers. Columns four and five show that average size is negatively correlated with markups as measured by either the average markup or the 90%-quantile of the within region-industry. This is consistent with the model if regional firm size is driven by differences in market access but not consistent with an explanation based on entry costs.

While these patterns are consistent with market barriers potentially playing an important role in the determination of misallocation, firm dynamics and growth, they are of course only suggestive. They however highlight the need to directly measure why the costs of entering products markets within countries might be systematically related to the level of development and whether these are amenable by policy.
4 Conclusion

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

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

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

As an application, I calibrate the theory to firm-level panel data in Indonesia. This setting is motivated by the recent literature on misallocation in developing countries. This literature typically takes the extent of misallocation as given and relies on exogenous firm-specific wedges as a modeling device. In the model presented in this paper, such wedges exactly coincide with firms’ markups and are therefore endogenous determined. I find that markups can plausibly account for 15% of the measured dispersion in average products and would reduce aggregate TFP by roughly 1%. The static efficiency losses from markups alone are therefore unlikely to be main culprit of aggregate TFP differences across countries. I also find that relative to the US benchmark, firms in Indonesia are particularly hurt by market barriers, i.e. frictions for existing firms to enter new product markets. In contrast, higher entry costs for new firms are less important. Quantitatively, such frictions increase of misallocation by 0.3%. Higher market barriers in Indonesia nevertheless 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


Appendix

A-1 Appendix: Theoretical Results

A-1.1 Proof of Proposition 1

I prove Proposition 1 in two steps. I first derive the value function (13). Then I show that the conditions in Proposition 1 uniquely define the optimal choices for \((I, x, z)\).

A-1.1.1 Solving for the value function \(V_t(\cdot)\) in (13)

To derive (13), conjecture first that the value function takes an additive form

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

Equation (9) then implies that \(V_t^P\) and \(V_t^M\) are defined by the differential equations

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

and

\[
rV_t^P(n) - \dot{V}_t^P(n) = n \times \pi_t(1) + \sum_{i=1}^n \tau \left[ V_t^P(n-1) - V_t^P(n) \right] + \max_X \left\{ X \left[ V_t^P(n+1) + V_t^M(1) - V_t^P(n) \right] - c^X(X, n) w_t \right\}. \tag{A-2}\]

Now consider a steady state where both value functions grow at rate \(g\) and assume the cost functions in (10). Then we can write (A-2) as

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

Conjecture that

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

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

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

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

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

\[46\text{The analysis in this section only contains the most important steps. A detailed derivation is contained in Section OA-1.1 of the Online Appendix.}\]
so that $\kappa_t = \frac{1}{\rho + \tau}$ and $\alpha_t = \frac{Y_t - \frac{\xi - 1}{\varphi \xi} I^\xi w_t}{\rho + \tau}$. Hence, (A-3) yields

$$V_t^M(\Delta) = \frac{(\lambda - \Delta) Y_t + \lambda - \Delta \frac{\xi - 1}{\varphi \xi} I^\xi w_t}{\rho + \tau} = \frac{\pi_t(\Delta) - \pi_t(1) + (\zeta - 1) c_t(I, \Delta) w_t}{\rho + \tau}.$$  

Note also that this implies that

$$V_t^M(1) = \frac{(\zeta - 1) w_t \frac{1}{\varphi \xi} I^\xi}{\rho + \tau}. \quad \text{(A-5)}$$

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

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

The optimality condition for $x$ reads

$$v_t + V_t^M(1) = \frac{\xi}{\varphi \xi} x^\xi -1 w_t. \quad \text{(A-6)}$$

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

$$(\rho + \tau) v_t = \pi_t(1) + (\zeta - 1) \frac{1}{\varphi \xi} x^\xi w_t. \quad \text{(A-7)}$$

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

$$\frac{\xi}{\varphi \xi} x^\xi -1 \varpi = \varpi + \frac{V_t^M(1)}{w_t} = \varpi + \frac{(\zeta - 1) \frac{1}{\varphi} I^\xi}{\rho + \tau}$$

$$\varpi = \frac{\lambda - 1 \frac{Y_t}{w_t} + \frac{\xi - 1}{\varphi \xi} x^\xi}{\rho + \tau}.$$  

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

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

Substituting into (A-1) yields (13).

**A-1.1.2 Existence and Uniqueness**

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

$$L_t^I = \int_{j=0}^{1} c^I(I, \Delta) dj = \int_{j=0}^{1} \frac{1}{\varphi \xi} \times I^\xi \lambda^{-\Delta} dj = \frac{1}{\varphi \xi} \times I^\xi \times \Lambda \left( \frac{\tau}{I} \right).$$
Hence, the equilibrium is defined by the four equations

\[ 1 = \Lambda \left( \frac{\tau}{I} \right) \times \frac{Y_t}{w_t} + \frac{1}{\varphi_I} \times I^\xi \times \Lambda \left( \frac{\tau}{I} \right) + \frac{1}{\varphi_I} \times (\tau - x) + \frac{1}{\varphi_x} \times x^\xi \]  \hspace{1cm} (A-8)

\[ \frac{1}{\varphi_z} = \frac{\lambda - 1}{\lambda} \frac{Y_t}{w_t} + \frac{\zeta - 1}{\varphi_I} x^\xi + \frac{\zeta - 1}{\varphi_I} \lambda I^\xi \]  \hspace{1cm} (A-9)

\[ \frac{Y_t}{w_t} = \frac{\lambda}{\lambda - 1} \frac{\zeta}{\varphi_I} I^{\xi - 1} (\rho + \tau) + (\zeta - 1) \frac{1}{\varphi_I} I^\xi \]  \hspace{1cm} (A-10)

\[ Y_t = \frac{\zeta}{\varphi_I} x^{\xi - 1} (\rho + \tau) \frac{\lambda}{\lambda - 1} - \frac{\lambda}{\lambda - 1} \frac{\zeta - 1}{\varphi_I} x^\xi - \frac{1}{\varphi_I} \frac{\zeta - 1}{\varphi_I} I^\xi. \]  \hspace{1cm} (A-11)

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

\[ x^{\xi - 1} = \frac{\varphi_x}{\varphi_z} \zeta. \]  \hspace{1cm} (A-12)

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

\[ 1 = \Lambda \left( \frac{\tau}{I} \right) \times \left( \frac{\lambda}{\lambda - 1} \frac{\zeta}{\varphi_I} I^{\xi - 1} (\rho + \tau) + \frac{\zeta - 1}{\varphi_I} I^\xi \right) + \frac{1}{\varphi_I} \times \tau - h(\phi). \]  \hspace{1cm} (A-13)

\[ \frac{1}{\varphi_z} = \frac{\zeta}{\varphi_I} I^{\xi - 1} + \frac{\zeta - 1}{\varphi_I} \frac{I^\xi}{\rho + \tau} + \frac{h(\phi)}{\rho + \tau}, \]  \hspace{1cm} (A-14)

where

\[ h(\phi) = \left( \frac{\zeta - 1}{\zeta} \right) \left( \frac{\varphi_x}{\zeta \varphi_z} \right) - \zeta \geq 0. \]  \hspace{1cm} (A-15)

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

\[ 1 = \Lambda(s) \times \left( \frac{\lambda}{\lambda - 1} \frac{\zeta}{\varphi_I} I^{\xi - 1} (\rho + sI) + \frac{\zeta - 1}{\varphi_I} I^\xi \right) + \frac{1}{\varphi_I} \times sI - h(\phi) \]  \hspace{1cm} (A-16)

\[ \frac{1}{\varphi_z} = \frac{\zeta}{\varphi_I} I^{\xi - 1} + \frac{\zeta - 1}{\varphi_I} \frac{I^\xi}{\rho + sI} + \frac{h(\phi)}{\rho + sI}, \]  \hspace{1cm} (A-17)

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

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

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

\[ \frac{1}{\varphi_z} = \frac{\zeta}{\varphi_I} I^{\xi - 1} + \frac{1}{\rho + sI} \left( \frac{\zeta - 1}{\varphi_I} I^\xi + h(\phi) \right) = G \left( I^{FE}(s), s \right). \]  \hspace{1cm} (A-18)
Clearly, $\frac{\partial G(I,s)}{\partial s} < 0$. Also
\[
\frac{\partial G}{\partial I} = \frac{(\zeta - 1) \zeta I^{\zeta - 1} \frac{1}{\zeta} - (\rho + sI)^{-2} s \left[ (\zeta - 1) \frac{1}{\zeta} I^{\zeta} + h(\varphi) \right] + (\rho + sI)^{-1} (\zeta - 1) \frac{1}{\zeta} I^{\zeta - 1}}{\zeta - 1}.
\]

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

As $I^{FE}(0) > 0$, this requires that $\frac{1}{\varphi} > \frac{h(\varphi)}{\rho}$. From (A-15) we can write this condition as
\[
\rho > \varphi \left( \frac{\zeta - 1}{\zeta} \right) \frac{1}{\varphi z} \frac{1}{\zeta} = \left( \frac{\zeta - 1}{\zeta} \right) \frac{1}{\zeta z} \frac{1}{\zeta} . \tag{A-19}
\]

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

### A-1.2 Proof of Proposition 2

Consider first the distribution of quality gaps $\nu(\Delta, t)$. In a stationary equilibrium we have $\dot{\nu}(\Delta, t) = 0$. (16) then implies that
\[
\nu(\Delta) = \left( \frac{I}{I + \theta} \right)^{\Delta} \frac{\Delta}{\zeta} \frac{\Delta}{I} = \left( \frac{1}{I + \theta} \right)^{\Delta} \frac{\Delta}{\zeta} = \left( \frac{1}{I + \theta} \right)^{\Delta} \frac{\Delta}{\zeta} .
\]

Hence, $P(\Delta \leq \delta) = 1 - \left( \frac{1}{I + \theta} \right)^{\delta} = 1 - e^{-\theta(1+\theta)\times d}$. This implies that log markups in $\mu = \Delta \ln(\lambda)$ are exponentially distributed with parameter $\theta$. Similarly, $F(\mu; x) = P(\lambda \Delta \leq \mu) = 1 - e^{-\mu/\theta}$. To derive (17), note that
\[
\Lambda = \int \mu - \theta \mu^{-\theta + 1} d\mu = \frac{\theta}{\theta + 1},
\]
\[
M = \exp(-E \ln(\mu)) \Lambda^{-1} = e^{-\theta - 1} \frac{\theta + 1}{\theta} .
\]

### A-1.3 Proof of Proposition 3

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

**The distribution of markups as a function of product age: equation (18)** I first show that the distribution of quality gaps $\Delta$ as a function of age conditional on survival, $\zeta(\Delta) \mu$, is given by $\zeta(\Delta) \mu = \frac{1}{\lambda} \left( Ia \right) \lambda e^{-Ia}$. Let $p(\Delta) a$ denote the probability of the product having a quality gap $\Delta$ at age $a$ when it was introduced at time 0. The corresponding flow
\[
\Lambda^{\Delta} = \sum_{i=1}^{\infty} \lambda^{-i} \mu(i) = \frac{e^x}{\lambda(\lambda + 1)} \sum_{i=1}^{\infty} \left( \frac{1}{\lambda(\lambda + 1)} \right)^i = \frac{\tau}{\lambda + \lambda - 1} .
\]

Similarly, $\int_{0}^{\Delta} \Delta(v) d\mu = \sum_{i=1}^{\infty} i \mu(i) = \frac{\tau}{\lambda} \sum_{i=1}^{\infty} i \left( \frac{1}{\lambda(\lambda + 1)} \right)^i = \frac{\tau}{\lambda} \sum_{i=1}^{\infty} i \left( \frac{1}{\lambda(\lambda + 1)} \right)^i = 1 + 1 \frac{\tau}{\lambda} \sum_{i=1}^{\infty} i \left( \frac{1}{\lambda(\lambda + 1)} \right)^i$, so that $M^{\Delta} = \frac{1}{\lambda(\lambda + 1)} \lambda^{2} \frac{\tau}{\lambda} \sum_{i=1}^{\infty} i \left( \frac{1}{\lambda(\lambda + 1)} \right)^i$. 

A-4
equation are

\[ p_\Delta (a) = \begin{cases} (1 - p_0 (a)) \tau & \text{for } \Delta = 0 \\
-p_1 (a) (I + \tau) & \text{for } \Delta = 1 \\
p_{\Delta-1} (a) I - p_\Delta (a) (I + \tau) & \text{for } \Delta \geq 2 \end{cases} \]

The solution to this set of differential equations is given by

\[ p_0 (a) = 1 - e^{-\tau \times a} \]
\[ p_i+1 (a) = \left( \frac{1}{i!} \right) I^i (e^{-\tau \times a}) \text{ for } i \geq 0. \]

The distribution of markups conditional on survival is then

\[ p_{S,i+1} (a) \equiv p_{i+1} (a) \left( 1 - p_0 (a) \right) = \left( \frac{1}{i!} \right) I^i (e^{-\tau \times a}) \text{ for } i \geq 0. \]

This is a Poisson distribution with parameter \( Ia \), so that

\[ E[\Delta \mid a] = Ia. \quad (19) \]

then follows because \( \ln (\mu) = \ln (\lambda) \Delta. \)

The expected log markup by age: equation (20) From (3) we know that firm-level markups are given by

\[ \mu_f = \frac{\mu_f}{wl_f} = \frac{1}{\sum_{j=1}^{n} \lambda^{-\Delta_j}}. \]

Hence,

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

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

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

Then

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

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

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

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

Let $U$ denote the age of the product so that

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

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

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

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

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

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

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

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

$$E[\Delta|a, n, ni] = 1 + I \times \frac{\frac{1}{2} (1 - e^{-\tau a}) - \frac{1}{2} e^{-\tau a} (1 - \exp(-xa))}{1 - \exp(-(x+\tau)a)}. \quad (A-24)$$

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

$$E[\Delta|a, n, i] = 1 + Ia. \quad (A-25)$$

**Solving for $P(B = 0|a, n)$**  

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

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

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

$$H_Q(z, t) \equiv \sum_{n=0}^{\infty} Q(n, t) z^n. \quad (A-27)$$
Then \( \frac{\partial H_Q(z,t)}{\partial z} = \sum_{n=1}^{\infty} nQ(n,t)z^{n-1} \) and \( \frac{\partial H_Q(z,t)}{\partial t} = \dot{Q}(0,t) + \sum_{n=1}^{\infty} \dot{Q}(n,t)z^n \). Using (A-26) it follows that,

\[
\frac{\partial H_Q(z,t)}{\partial t} = \tau p(1,t) + \sum_{n=1}^{\infty} [x(n-1) \times Q(n-1,t) + \tau (n+1) \times Q(n+1,t) - n(x+\tau)Q(n,t)]z^n 
+ \tau \sum_{n=1}^{\infty} p(n+1,t)z^n - \tau \sum_{n=1}^{\infty} Q(n+1,t)z^n 
= \frac{\tau}{z} (H_P(z,t) - H_Q(z,t)) + (xz^2 - (x+\tau)z + \tau) \frac{\partial H_Q(z,t)}{\partial z},
\]

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

\[
H_P(z,t) = \sum_{n=0}^{\infty} p(n,t)z^n, \quad \text{(A-28)}
\]

Now define

\[
\Psi(z,t) = H_P(z,t) - H_Q(z,t). \quad \text{(A-29)}
\]

Then

\[
\dot{\Psi}(z,t) = \dot{H}_P(z,t) - \dot{H}_Q(z,t) = (xz^2 - (x+\tau)z + \tau) \frac{\partial \Psi(z,t)}{\partial z} - \frac{\tau}{z} \Psi(z,t), \quad \text{(A-30)}
\]

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

\[
\Psi(z,0) = \sum_{n=0}^{\infty} p(n,0)z^n - \sum_{n=0}^{\infty} Q(n,0)z^n = z, \quad \text{(A-31)}
\]

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

\[
\Psi(z,t) = \frac{(\tau - x)z \times e^{-\tau t}}{x(z - 1) \times e^{-x(t-z)}t - (xz - \tau)}. \quad \text{(A-32)}
\]

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

\[
H_P(z,t) = \frac{\tau (z - 1) \times e^{-x(t-z)}t - (xz - \tau)}{x(z - 1) \times e^{-x(t-z)}t - (xz - \tau)}. \quad \text{(A-33)}
\]

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

\[
H_Q(z,t) = \Psi(z,t) - H_P(z,t) = \frac{\tau (z - 1) \times e^{-x(t-z)}t - (xz - \tau) - (z - x)z \times e^{-\tau t}}{x(z - 1) \times e^{-x(t-z)}t - (xz - \tau)}. \quad \text{(A-34)}
\]

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

\[
Q(n,t) = \left(1 - \frac{\tau e^{-xt} - x e^{-\tau t}}{\tau - x}\right) \times p(n,t), \quad \text{(A-35)}
\]
where \( p(n,t) \) is described by \( p(0,t) = \frac{z}{x} \gamma(t), \) \( p(1,t) = (1 - \gamma(t))(1 - p(0,t)) \) and \( p(n,t) = \gamma(t)^{n-1} p(1,t) \) and the function \( \gamma(t) \) is given by

\[
\gamma(t) = \frac{x(1 - e^{-(\tau-x)t})}{\tau - x \times e^{-(\tau-x)t}}.
\]

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

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

Hence,

\[
1 - P(B = 0|a,n) = \frac{\tau e^{-xa} - xe^{-\tau a}}{\tau - x}.
\]  

(A-35)

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

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

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

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

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

where

\[
E[a_P] = \frac{\frac{1}{x}(1 - e^{-\tau a}) - \frac{1}{x}e^{-\tau a} (1 - e^{-xa})}{1 - e^{-(\tau+x)a}} + \left( a - \frac{\frac{1}{x}(1 - e^{-\tau a}) - \frac{1}{x}e^{-\tau a} (1 - e^{-xa})}{1 - e^{-(\tau+x)a}} \right) \times \left( \frac{\tau e^{-xa} - xe^{-\tau a}}{x(1 - e^{-(\tau-x)a})} \right) \times \ln \left( \frac{\tau - x \times e^{-(\tau-x)a}}{\tau - x} \right).
\]

This is the required expression in \((20)\).

### A.1.4 Proofs for Section 2.7

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

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

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

\[
F = \frac{z}{\tau} \times \sum_{j=1}^{\infty} \frac{1}{j} \left( \frac{x}{\tau} \right)^{j-1} = \frac{z}{x} \times \sum_{j=1}^{\infty} \frac{1}{j} \left( \frac{x}{\tau} \right)^{j} = \frac{z}{x} \times \ln \left( \frac{z + x}{z} \right) = \frac{1 - \vartheta_x}{\vartheta_x} \ln \left( \frac{1}{1 - \vartheta_x} \right).
\]
The share of products produced by firms with at most $k$ products is given by

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

To derive the employment life cycle, i.e. equation (22), consider first the distribution of sales conditional on age. Note that $E[\ln n|a] = E \left[ \ln \left( \frac{n}{\ln n} \right) \right]|a = \ln \left( \frac{x}{\tau} \right) + E[\ln n|a] - E[\ln \mu_f|a]$. To calculate $E[\ln n|a]$ note that the distribution of $n$ conditional on age is given by $f_{N|A}(n|a) = \gamma(a)^{n-1} \times (1 - \gamma(a))$. Hence,

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

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

### A-1.5 Proof of Proposition 4

I prove the different parts in turn.

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

$$ 1 = \frac{\tau}{\lambda \tau + I(\lambda - 1)} \times \left( \frac{\lambda}{\lambda - 1} - \frac{\zeta}{\varphi I} I^\zeta - 1 (\rho + \tau) + \frac{\zeta}{\varphi I} I^\zeta \right) + \frac{1}{\varphi z} \times \tau - h(\varphi) = H^\tau(I, \tau) - h(\varphi|A-36) $$

$$ \frac{1}{\varphi z} = \frac{\zeta}{\varphi I} I^\zeta + \frac{\zeta}{\varphi I} I^\zeta \frac{\zeta - 1}{\rho + \tau} + \frac{h(\varphi)}{\rho + \tau} = G^\tau(I, \tau). \quad (A-37) $$

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

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

The numerator is positive for $\zeta \geq 2$ and negative for $\zeta = 1$. Hence, that is some $\zeta \geq \zeta$ such that $\frac{\partial H^\tau}{\partial \tau} > 0$. A sufficient condition is $\zeta \geq 2$. Hence, (A-36) implies a schedule $I^H(\tau)$, which decreasing and (A-37) implies a schedule $I^G(\tau)$, which is increasing. Furthermore, note that (A-15) implies that

$$ \frac{\partial h(\varphi)}{\partial \varphi z} = \frac{\zeta}{\zeta - 1} h(\varphi) > 0 $$

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

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

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

3. From (14) it is immediate that higher entry costs increase $x$. Hence, higher entry costs increase $\vartheta z$. 
4. Use (A-15) to write \( \tau = \frac{\zeta}{\zeta - 1} x \times x = \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi) \). Then we can write (A-36) and (A-37) as

\[
1 = \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi) \times \left( \lambda - 1 \frac{\zeta}{\varphi_1} I_{\varphi}^{-1} \left( \rho + \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi) \right) + \frac{\zeta}{\varphi_1} I_{\varphi} \right) + \left( \frac{1}{\sigma_x \zeta - 1} - 1 \right) h(\varphi)
\]

\[
\frac{1}{\varphi_z} = \frac{\zeta}{\varphi_1} I_{\varphi}^{-1} + \frac{\zeta - 1}{\varphi_1} I_{\varphi} + \frac{h(\varphi)}{\rho + \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi)} + \frac{h(\varphi)}{\rho + \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi)}.
\]

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

\[
\frac{\partial G}{\partial h} = \frac{\rho + \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi) - \left( \frac{\zeta - 1}{\varphi_1} I_{\varphi} + h(\varphi) \right) \frac{1}{\sigma_x \zeta - 1} \varphi_z}{\left( \rho + \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi) \right)^2} = \frac{\frac{\zeta - 1}{\varphi_1} I_{\varphi} \frac{1}{\sigma_x \zeta - 1} \varphi_z}{\left( \rho + \frac{1}{\sigma_x \zeta - 1} \varphi_z h(\varphi) \right)^2}.
\]

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

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

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

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

The Equilibrium Value Function and the Equilibrium Conditions

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

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

where

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

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

\[
v_t = \frac{1 - \frac{1}{\lambda}}{\lambda} Y_t + (\zeta - 1) \frac{1}{\varphi_1 \phi^z} x^z w_t
\]

\[
\rho + \tau
\]
The optimal innovation and expansion rates are given by
\[
I = \left( E \left[ 1 - \lambda^{-j} \right] \frac{V}{w_0} \frac{Y_t}{w_t} - (\zeta - 1) \frac{1}{\varphi_x} I^\zeta \right) \frac{1}{\varphi_x}. 
\]

\[
x = \left( \frac{\varphi_x}{\varphi_x} \right) \frac{1}{\varphi_x}. 
\]

The free entry condition is given by
\[
\frac{1}{\varphi_e} = \sum_{j=1}^{\infty} \left( \frac{V_t^P (1) + V_t^M (j)}{w_t} \right) p_j = \frac{1}{\rho + \tau} \left( \frac{Y_t}{w_t} + (\zeta - 1) \frac{1}{\varphi_x} x^\zeta + (\zeta - 1) \frac{1}{\varphi_x} I^\zeta - \frac{Y_t}{w_t} \right) \sum_j \lambda^{-j} p_j. 
\]

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

The Distribution of Markups and First-Order Stochastic Dominance

The distribution of quality gaps \( \nu (\Delta) \) solves the set of differential equations
\[
\hat{\nu} (\Delta, t) = \begin{cases} 
- (\tau + I) \nu (\Delta, t) + I \sum_{j=1}^{\Delta-1} \nu (\Delta - j, t) p_j + \tau p_\Delta & \text{if } \Delta \geq 2 \\
\tau (p_1 - \nu (1, t)) - \nu (1, t) I & \text{if } \Delta = 1 
\end{cases}.
\]

The stationary distribution is therefore given by
\[
\nu (j) = \frac{1}{1 + \vartheta} \left( \sum_{m=1}^{j-1} \nu (m) p_{j-m} \right) + \frac{\vartheta j}{1 + \vartheta} p_j, \quad (A-38)
\]

Define the cdf of quality gaps and hence markups as \( \Phi (k) = \sum_{j=1}^{k} \nu (j) \). I now show that
\[
\vartheta_H > \vartheta_L \Rightarrow \Phi (k; \vartheta_H) > \Phi (k; \vartheta_L) \text{ for all } k.
\]

To see this, define \( \alpha = \frac{\vartheta}{1 + \vartheta} \). \( \alpha \) is increasing in \( \vartheta \). Write \((OA-24)\) as
\[
\nu (j) = (1 - \alpha) \left( \sum_{m=1}^{j-1} \nu (m) p_{j-m} \right) + \alpha p_j.
\]

The cdf \( \Phi \) can be written as
\[
\Phi (k) = \sum_{j=1}^{k} \nu (j) = (1 - \alpha) \sum_{j=1}^{k} \sum_{m=1}^{j-1} \nu (m) p_{j-m} + \alpha \sum_{j=1}^{k} p_j = (1 - \alpha) \sum_{m=1}^{k-1} p_{k-m} \Phi (m) + \alpha \sum_{j=1}^{k} p_j.
\]

Let \( \Phi (k; \alpha_H) \) denote the cdf as a function of \( \alpha \). Then
\[
\Phi (k; \alpha_H) - \Phi (k; \alpha_L) = (1 - \alpha_H) \sum_{m=1}^{k-1} p_m \Phi (k - m; \alpha_H) + \alpha_H \sum_{j=1}^{k} p_j - (1 - \alpha_L) \sum_{m=1}^{k-1} p_m \Phi (k - m; \alpha_L) - \alpha_L \sum_{j=1}^{k} p_j
\]
\[
= (1 - \alpha_H) \sum_{m=1}^{k-1} p_{k-m} [\Phi (m; \alpha_H) - \Phi (m; \alpha_L)] + (\alpha_H - \alpha_L) \left[ p_k + \sum_{j=1}^{k} p_j (1 - \Phi (k - j; \alpha_L)) \right]. \quad (A-39)
\]
Now note that
\[ \Phi(1; \alpha_H) - \Phi(1; \alpha_L) = (\alpha_H - \alpha_L) p_1 > 0. \]

Furthermore, (A-39) implies that
\[ \Phi(m; \alpha_H) - \Phi(m; \alpha_L) > 0 \quad \text{for all } m < j \rightarrow \Phi(j; \alpha_H) - \Phi(j; \alpha_L) > 0 \]
as \( 1 - \Phi(k - j; \alpha_L) > 0 \) by \( \Phi \) being a cdf. This shows that \( \Phi(m; \alpha_H) - \Phi(m; \alpha_L) > 0 \) for all \( m \).

**The Case of** \( p_n = \frac{1-k}{k} \times \kappa^n \)

Suppose the step size is drawn from
\[ p_n = \frac{1-k}{k} \times \kappa^n. \tag{A-40} \]
This distribution is parametrized by a single parameter \( \kappa \). For \( \kappa \to 0 \), we recover the baseline model.

We now first show that the distribution of markups is again a pareto distribution. Using (OA-24) and (A-40), the density \( \nu_j \) solves the equation
\[ \nu_j = (1 - \alpha) \left( \sum_{m=1}^{j-1} \nu_m \frac{1-k}{k} \kappa^{j-m} \right) + \alpha \frac{1-k}{k} \kappa^j, \]
where \( \alpha = \frac{\vartheta}{1+\vartheta} \). Conjecture that \( \nu_j = A^{j-1} \nu_1 \) for \( j \geq 2 \). Substituting above yields
\[ A^{j-1} \nu_1 = \left( 1 - \alpha \right) \frac{1-k}{k} \kappa^{j-1} \left( 1 + \sum_{m=1}^{j-2} \left( \frac{A}{k} \right)^m \right) \kappa^{j-1} \nu_1. \]

It is easy to show that \( A = 1 - (1-k) \alpha = \frac{1+k \vartheta}{1+\vartheta} \) solves this equation. Note that \( \nu_1 = \frac{\vartheta}{1+\vartheta} p_1 = \frac{(1-k)\vartheta}{1+\vartheta} \). Hence,
\[ \nu_j = \left( \frac{1+k \vartheta}{1+\vartheta} \right)^j \frac{\vartheta(1-k)}{1+k \vartheta} \nu_1. \]

The corresponding cdf is given by
\[ \Phi(k) = \sum_{m=1}^{k} \nu_m = \nu_1 \sum_{m=1}^{k} A^{m-1} = \nu_1 \sum_{m=0}^{k-1} A^m = \nu_1 \frac{1-A^k}{1-A} = 1 - \left( \frac{1+k \vartheta}{1+\vartheta} \right)^k. \]

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

The distribution of markups is given by
\[ P[\mu \leq m] = P[\lambda^\Delta \leq m] = P[\Delta \leq \frac{\ln m}{\ln \lambda}] = 1 - e^{\frac{\ln m}{\ln \lambda} \times \ln \left( \frac{1+k \vartheta}{1+\vartheta} \right)} = 1 - m^{-\frac{1+k \vartheta}{\ln \lambda} \ln \left( \frac{1+k \vartheta}{1+\vartheta} \right)}. \]

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

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

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

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

The static allocations summarized in (23), (24) and (25) can be derived by standard argument. The dynamic environment for own-innovation, entry and incumbent creative destruction is the same as in the baseline model. The only difference with respect to the baseline is that I assume that a fraction $(1 - \delta)$ of creative destruction activities result in a “reset” of the quality of the destroyed product to the level $\lambda Q_t$. This change is necessary to make the productivity distribution stationary. I discuss this in much more detail in Section OA-1.5 in the Online Appendix. All the aggregate implications independent of the parameter $\delta$. The need for a stationary distribution of quality only arises when taking the model to the data. For continuity with the baseline model I still assume that the quality gap $\Delta$ after such a reset is still equal to unity.

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

$$r_t V_t ([\Delta_i, q_i]_{i=1}^{n}) - \dot{V}_t ([\Delta_i, q_i]_{i=1}^{n}) = \sum_{i=1}^{n} \pi_t ([\Delta_i, q_i]_{i=1}^{n}) + \sum_{i=1}^{n} r_t \left[ V_t ([\Delta_j, q_j]_{j \neq i}) - V_t ([\Delta_i, q_i]_{i=1}^{n}) \right] + \max_{[I_t]_{t=1}^{\infty}} \left\{ \sum_{i=1}^{n} I_t \left[ V_t ([\Delta_i, q_i]_{i=1}^{n}, [\Delta_i + 1, \lambda q_i]) - V_t ([\Delta_i, q_i]_{i=1}^{n}) \right] \right\} + \max_{X} \left\{ X \left[ \delta \int q V_t ([\Delta_i, q_i]_{i=1}^{n}, 1, \lambda q) dF_t (q) + (1 - \delta) \sum_{i=1}^{n} c^I (I_t, q_i) w_t \right] \right\} + \max_{X} \left\{ X \left[ \delta \int q V_t ([\Delta_i, q_i]_{i=1}^{n}, 1, \lambda q) dF_t (q) + (1 - \delta) \sum_{i=1}^{n} c^I (I_t, q_i) w_t \right] \right\} - c^X (X, n [\Delta_t]_{t=1}^{\infty}) \right\}$$

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

$$\frac{1}{\bar{\varphi}_z} w_t = \delta \int q V_t (1, \lambda q) dF_t (q) + (1 - \delta) V_t (1, \lambda Q_t).$$

(A-43)

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

$$c^I_t (I; \Delta, q) = \frac{1}{\varphi_t} \left( \frac{q}{Q_t} \right)^{\kappa - 1} I^\kappa.$$

(A-44)

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

$$V_t ([\Delta_i, q_i]_{i=1}^{n}) = \sum_{i=1}^{n} \frac{\psi (\Delta_i)}{\rho + \tau + (\sigma - 1) g} \left( \frac{q_i}{Q_t} \right)^{\kappa - 1} \frac{Y_t}{E [\mu^{\kappa - \sigma}]} + \frac{1}{\rho + \tau} \frac{\kappa - 1}{\varphi_x} \left( \frac{\varphi_x}{\varphi_z} \right)^{\frac{1}{\kappa - \sigma}} w_i n,$$
where $\psi(\Delta_i)$ is implicitly defined and depends only on $\Delta$ (and general equilibrium variables). I also show that the optimal innovation rate is given by

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

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

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

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

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

(A-45)

Note that if $I(j) = I$ we have $\nu(\Delta) = \frac{\tau}{I(1)} \left( \frac{1}{\tau + I(1)} \right)^{\Delta} = \partial_1 \left( \frac{1}{\tau + I(1)} \right)^{\Delta}$ as in the baseline model. Given that markups are a one-to-one function of quality gaps (see (OA-25)), the distribution of markups is also stationary and only a function of $\tau$ and $\{I(\Delta)\}_{\Delta=1}^{\infty}$.

A-2 Appendix: Empirical Results

A-2.1 Measuring markups

To measure markups I closely follow the approach of De Loecker and Warzynski (2012). The crucial empirical object to implement (30) is the firms’ labor share $s_{l,ft} = \frac{w_{l,ft}}{p_{l,ft} v_{a,ft}}$. As pointed out by De Loecker and Warzynski (2012), the level of production $y_{ft}$ might contain both unanticipated shocks to and measurement error. Hence, they propose to consider a regression of

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

(A-46)

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

$$s_{l,ft} = \frac{w_{l,ft}}{v_{a,ft}/\exp(\hat{e}_{ft})},$$

where $v_{a,ft}$ is observed value added, $\hat{e}_{ft}$ is the residual from (A-46), where I take $v_{a,ft}$ is the dependent variable and take $\phi(.)$ a second order polynomial in all (log) inputs and their interaction terms. For the specification with
Table A-1: The employment life cycle in Indonesia

intermediate inputs instead of labor, the procedure is analogous.

A-2.2 The employment life cycle in Indonesia

The calibration uses the employment life cycle in Indonesia as an explicit calibration target. Focusing on the unbalanced panel of firms entering the economy after 1991, firms grow by about 8% a year. For the calibration I therefore use the estimated profile depicted in Figure 4 and target the log difference in employment for 7 year old firms. In Table A-1 I report additional regression results of predicting firm employment from firm age. The specification is exactly the same as (31), except that I do not control for firms' capital intensity. Column 3 shows that entrants and exiting firms are substantially smaller than the average firm. Column 4 shows that entrants are not too small given their age (in fact, they are slightly bigger) but that exiting firms are much smaller. This is exactly what the model predicts, because exiting firms are selected on \( n \) conditional on age, while entrants are not. Column 5 shows that the estimated age profile is quite similar once I condition on survival until the end of the sample. Finally, the last column controls for firm fixed effects. This lowers the age coefficient substantially.

A-2.3 The markup-size relationship

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

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

Above I already showed that

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

$$E[\ln(\mu_f)|N = n] = \int_a \left[ E[\Delta|a,ni] + \frac{1 - P(B = 0|a)}{n} \left( E[\Delta|a,i] - E[\Delta|a,ni] \right) \right] f(a|n) da$$

$$= 1 + I \times \int_a \left[ g(a,x,\tau) + \frac{(1 - P(B = 0|a))}{n} \left( a - g(a,x,\tau) \right) \right] f(a|n) da \quad (A-47)$$

where

$$g(a,x,\tau) = \frac{\frac{1}{2} (1 - e^{-\tau a}) - \frac{1}{x} e^{-\tau a} (1 - \exp(-xa))}{1 - \exp(-(x + \tau) a)},$$

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

$$f(a|n) = \frac{(1 - \gamma(a)) \gamma(a)^{n-1} \left( 1 - \frac{\tau}{x} \gamma(a) \right)}{\frac{1}{x} \gamma(a)} \frac{\frac{1}{x} \left( \frac{x}{\gamma(a)} \right)^n}{\frac{1}{x} \left( \frac{x}{\gamma(a)} \right)^n}, \quad (A-48)$$

where $\gamma(a) = \gamma(a) = \frac{1 - e^{-\tau a}}{1 - xe^{-\tau a}}$.

The expressions above fully determine the average log markup as a function of $n$. In Figure 8 I show the results for the calibrated model. The left panel shows the average markup as a function of size, i.e. (A-47). In the right panel I show the stationary firm size distribution.

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

---

48 To derive (A-48), note first that the mass of firms with $n$ products is given by

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

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

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

Each period $z$ firms enter. Hence,

$$M_n = \int_{a=0}^{\infty} z p_n(a) da.$$

Then conditional distribution is therefore given by

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

which is the expression in (A-48).
Notes: In the left panel I depict the average (log) markup as a function of ln n. In the right panel I depict the stationary distribution of sales. The endogenous flow rates (I, x, z) and the parameter λ correspond to the calibration of the baseline model.

Figure 8: Markups by Size

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>log Markups (ln µf)</th>
<th>log TFPR (ln py/(kα l1−α))</th>
</tr>
</thead>
<tbody>
<tr>
<td>log sales</td>
<td>0.192***</td>
<td>0.168***</td>
</tr>
<tr>
<td></td>
<td>(0.000925)</td>
<td>(0.00109)</td>
</tr>
<tr>
<td>ln k/l</td>
<td>0.0547***</td>
<td>-0.242***</td>
</tr>
<tr>
<td></td>
<td>(0.00134)</td>
<td>(0.00137)</td>
</tr>
<tr>
<td>N</td>
<td>176958</td>
<td>138953</td>
</tr>
<tr>
<td></td>
<td>122578</td>
<td>122578</td>
</tr>
<tr>
<td>R²</td>
<td>0.306</td>
<td>0.311</td>
</tr>
<tr>
<td></td>
<td>0.153</td>
<td>0.354</td>
</tr>
</tbody>
</table>

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