Innovation, Knowledge Diffusion, and Selection

Danial Lashkari*

February 14, 2018

This paper constructs a theory of industry growth through innovation and selection-driven creative destruction. Firms' ideas determine their productivity and stochastically evolve over time. Firms innovate to improve their ideas and endogenously exit if unsuccessful. Entrants adopt the ideas of incumbents. In this model, when better ideas are innovated or adopted, they selectively replace worse ideas. Innovation externalities vary based on firm productivity: ideas generated by more productive firms create 1) longer-lasting positive externalities due to knowledge diffusion and 2) stronger negative externalities due to the displacement of other firms. Therefore, the net externalities of innovation are heterogeneous, and market equilibrium misallocates innovation across firms. Solving for the optimal allocation shows that the instruments of innovation policy should depend on firm productivity. A calibration of the model to firm-level data from US manufacturing and retail trade suggests quantitatively large misallocations that create first-order considerations for policy design.

*Boston College. Email: danial.lashkari@yale.edu. Website: https://sites.google.com/view/danial-lashkari/home. I am grateful for the advice and support of Pol Antràs, Robert Barro, Elhanan Helpman, and Marc Melitz. I have also immensely benefited from insightful conversations and the feedback from Costas Arkolakis, Jess Benhabib, Kirill Borusyak, Davin Chor, Diego Comin, Xavier Gabaix, Boyan Jovanovic, Rohan Kekre, Amir Kermani, Samuel Kortum, William Kerr, Martí Mestieri, Pooya Molavi, Michael Peters, Mikkel Plagborg-Møller, Francisco Queiró, Martin Rotemberg, and Peter Schott.

1 Introduction

Our canonical theories of economic growth champion firm-level innovation as the main engine of productivity growth. Inspired by this idea, over the past decades governments around the world have employed policies such as R&D tax incentives and research grants to promote and support innovation among private businesses.[1] These programs are typically justified as instruments to internalize the externalities that growth theories attribute to firm innovation. In particular, in models of creative destruction (Aghion and Howitt, 1992; Grossman and Helpman, 1990) innovating firms generate positive *knowledge spillovers* for other firms while negatively impacting the business of their product market rivals through *displacement effects* (e.g., business stealing). A burgeoning literature has recently begun to use this theoretical framework to examine the impact of innovation policies on aggregate productivity growth and to study their optimal design (Acemoglu et al., 2013; Lentz and Mortensen, 2014; Atkeson and Burstein, 2015; Akcigit et al., 2016).

In bringing theories of firm innovation to the design of innovation policy, accounting for firm heterogeneity has emerged as a key consideration.[2] Firm heterogeneity matters to the extent that innovation externalities vary across firms with different productivity and size (Atkeson and Burstein, 2015). The available empirical evidence suggests that the two types of innovation externalities discussed above may in fact be heterogeneous across firms (Bloom et al., 2013). If so, we have to examine the incidence of innovation policy across different firms, or even to employ policy instruments that differentiate among them. To analyze such situations, most prior work has relied on the workhorse model of Klette and Kortum (2004), modeling innovating firms as entrants that build on the knowledge of an incumbent and simultaneously displace it in the market. Despite its conceptual and analytical elegance, this head-to-head account of competition also creates some limitations. In particular, it mechanically ties together the two conceptually distinct positive and negative externalities of innovation: receiving knowledge spillovers from an incumbent firm always coincides with displacing that firm in the market.

This paper proposes an alternative formulation for the process of creative destruction that disentangles knowledge spillovers from displacement effects. Knowledge spillovers stem from the *diffusion* of productive ideas to new firms, while displacement effects are driven by the *selection* of unproductive ideas out of the industry.[3] In the model, entrants can adopt the ideas of productive

[1] For broad examples of these programs across developed and developing countries, see OECD (2012) and World Bank (2010). For more information on the design and scope of R&D tax incentives, see OECD (2010) and OECD (2011).

[2] When firms vary in productivity, the reallocation of market shares from less to more productive firms may also contribute to aggregate productivity growth (Syverson, 2011). Studies have found that, among these reallocations, the margin of firm entry and exit constitutes a particularly large component of productivity growth across different industries (e.g., Foster et al., 2001, 2006; Bartelsman and Haltiwanger, 2009). Therefore, to provide a quantitatively sound micro-macro link, our theories of firm-level innovation have to account for the contribution of these reallocations, both in the intensive and extensive margins, to aggregate growth.

[3] I use the broad term displacement effect, rather than business stealing, to accommodate multiple related externalities. Note that in standard models of creative destruction there are already two distinct negative externalities tied to firm displacement. The first one is the (static) business stealing effect whereby an innovating firm steals the profits of a displaced firm and negatively impacts its outcomes. The second one arises since the innovating firm also displaces the incumbent's ability to innovate in that market (see Lentz and Mortensen, 2014, for a clear illustration of this point). Crucially, the model presented in this paper will feature yet other displacement externalities despite the fact that the static business stealing inefficiencies are removed due to the CES assumption.

firms to displace unproductive firms in the market. I show that incorporating selection as the core of creative destruction unleashes a host of novel, economically meaningful, and sizable sources of heterogeneity in innovation externalities that matter for policy design. I use the model to derive policies that optimally reallocate innovation investments, and qualitatively and quantitatively examine their properties. Finally, I apply a calibration exercise to compare these policies across two broad sectors of the US economy, manufacturing and retail trade.

The model builds on the recent theories of knowledge diffusion and selection (Luttmer, 2007, 2012; Perla and Tonetti, 2014; Sampson, 2016) and extends them to include process innovation (Ericson and Pakes, 1995; Atkeson and Burstein, 2010). In the model, firms monopolistically compete in a differentiated product market à la Melitz (2003). Entrants receive knowledge spillovers from incumbent firms. Heterogeneous incumbents invest in innovation to raise their productivity and may endogenously exit due to fixed costs. Entry, exit, innovation, and the distribution of productivity in the industry are all endogenous. Three sets of structural features characterize these endogenous outcomes: costs of entry and innovation, volatility of firm productivity, and the strength of knowledge spillovers. I specify the strength of knowledge spillovers through two different characteristics of the process of entry and adoption: first, the imperfection of adoption, in the form of decreasing returns to adopting high productivity ideas, and second, directedness of adoption, in the sense of the probability of adopting the frontier idea in the industry (see Eeckhout and Jovanovic, 2002).

The interplay between diffusion and selection create novel sources of externalities in the model. Selection implies that the ideas of firms gradually become obsolete as better ideas emerge elsewhere in the industry. This gives rise to a new margin of heterogeneity across ideas: their expected lifetime in the market.[4] More productive ideas take longer before becoming obsolete. They therefore generate greater knowledge spillovers as, over the course of their longer lifetimes, more entrants have the opportunity to adopt them. Therefore, knowledge spillovers vary with the productivity of the ideas of firms. In addition, when a firm innovates, it intensifies future market competition, which then reduces the expected lifetimes of other firms and lowers their incentives for innovation. With each investment, therefore, a firm reallocates innovation investments across other firms and toward itself. Since the knowledge spillover externalities are heterogeneous, this reallocation produces a second external effect on aggregate productivity growth.

To see the intuition behind these new forces, consider the following example. Walmart, the largest and most productive firm in the US retail sector, has been responsible for many transformative technological and managerial innovations in this sector.[5] These innovations have fueled the expansion of Walmart, allowing it over time to replace many less productive local retailers. But, in doing so, Walmart has not necessarily built on the knowledge that had gone into the production processes of the displaced firms. Rather, Walmart has simply brought in alternative strategies for the organization of retail stores, inventory and supply management, and pricing that have made the practices of the displaced firms obsolete. In addition, Walmart ideas have diffused across the industry as new retail chains have adopted them to enter the market and displace other underperforming retailers. To the extent that Walmart innovations furnish such opportunities for

[4] Note that under the standard theories of creative destruction, the expected lifetime of all ideas are identical, since the likelihood of being displaced by another firm does not depend on the productivity of the idea.

[5] For an overview of these innovations and their impact on the retail industry, see (Basker, 2007).

new firms, Walmart innovations create positive externalities.[6] To the extent that they crowd out the innovations of the smaller firms that might have benefited other entrants, Walmart innovations create negative externalities. In designing policies that subsidize business investments to innovation, the government has to compare the magnitude of both two externalities for a large firm like Walmart with those for smaller and less productive retailers.

In addition to these novel normative aspects, the equilibrium of the model parsimoniously explains many stylized facts about firm heterogeneity and dispersion (Appendix 6 provides an overview of these facts).[7] One important analytical result is a decomposition of the rate of productivity growth into the direct contribution of incumbent innovation investments and the contribution of entry and exit (selection). The former is given by the market-share-weighted average of the investments of different incumbents. The latter is given by the product of the rate of firm turnover and the gap in the productivity of average entering and exiting firms. Due to spillovers to entrants and displacement effects on exiting firms, incumbent innovations indirectly affect the contribution of selection to growth. In particular, the gap in the productivity of industry entrants and exiting incumbents becomes endogenous in this model. This constitutes a new margin for the impact of firm innovation on aggregate growth.

I provide a partial equilibrium decomposition of innovation externalities into the positive knowledge spillover effects and the negative displacement effects. Typically, both knowledge spillover externalities and displacement effects grow in the productivity of the innovator. In general, the gaps between social and private returns to innovation may become a nonmonotonic function of productivity. More specifically, if the adoption of ideas is imperfect (involves decreasing returns), for highly productive firms the knowledge spillover effects eventually become dominated by the displacement effect. This is because the former grows more slowly than the market share of the innovator while the latter remains proportional to it.

Since externalities are heterogeneous under the market equilibrium, some firms overinvest and some underinvest in innovation. The social planner should therefore reallocate innovation investments across firms. I characterize the socially optimal allocation of innovation and production, and provide a method to compute it. Moreover, I derive the policy that implements this optimal allocation using taxes and subsidies to production and innovation. Under the assumption that we can observe the productivity ranking of firms, the derived policy uses a combination of a subsidy to startup costs for entrants, a lump-sum tax on the operation of firms, and a nonlinear subsidy to innovation investments that depends on the ranking of their productivity in the industry. The nonlinear innovation subsidy is the main instrument for the reallocation of innovation

[6] Case studies suggest that these forces are more than mere conceptual possibilities. For instance, in a study of the impact of Walmart on the US productivity growth in the late 90s, McKinsey Global Institute (2001) finds that the firm's innovations "directly and indirectly caused the bulk of the productivity acceleration through ongoing managerial innovation that increased *competition intensity* and drove the *diffusion* of best practice (both managerial and technological) [emphasis added]."

[7] Through the selection margin, the model readily predicts the robustly documented negative correlation between firm productivity and likelihood of exit (e.g., Foster et al., 2001). Moreover, volatility and selection together allow the model to endogenously generate a Pareto-tailed distribution of firm productivity and size, as is widely reported in empirical work on firm size distribution and industry concentration (Axtell, 2001; Luttmer, 2010). This feature is particularly important since key endogenous variables such as innovation investments and their external values all vary with firm productivity. Therefore, in order to explain the behavior of the industry both at the micro and at the aggregate levels, the model's prediction should fit the tail properties of the distribution of productivity.

investments.[8]

I examine the quantitative importance of these new forces by calibrating the model based on data from two broad sectors of the US economy, manufacturing and retail trade.[9] I rely on the Business Dynamics Statistics (BDS) data that tabulates moments of the life cycle dynamics of firm employment for each sector. I calibrate a total of 7 parameters, two of which are directly calibrated to the rates of productivity and employment growth in the sector. For the remaining five parameters, I use a simulated moment matching strategy. The identification of costs of innovation and volatility comes from the rates of employment growth and reallocation among mature firms. To identify the parameters of the process of entry, i.e., the imperfection and the directedness of adoption, I rely on the core logic of the model linking the quality of firm ideas to their expected lifetimes in the industry. According to this logic, any distribution on the initial productivity of a cohort of entrants translates into a distribution on the timing of their exit from the industry, conditional on the other model parameters. Therefore, I can use the empirically observed relationship between firm age and the hazard of exit to calibrate the parameters of the process of adoption. The results suggest that, despite its parsimony, the model provides a reasonable fit for the key moments in the data.

Investigating the schedule of subsidies that implement the optimal allocations, I find that for all but the very top firms we may approximate the policy by a subsidy rate that is an affine function of the productivity ranking of the firm. Incumbent firms receive a fixed subsidy rate for their innovation investments and a varying subsidy rate that is linear in their position relative to other firms in the distribution of productivity (with the rank normalized to be between 0 and 1). The dependence of the subsidy on the productivity ranking is weaker in retail compared to manufacturing (the approximate slope is about twice as large in manufacturing). Moreover, as productivity rises the pattern eventually reverses and the subsidy rate decreases among the most productive firms. This pattern is more pronounced in retail trade, due to the stronger decreasing returns to adoption implied by the calibrated parameters in this sector. For large firms in retail, their negative displacement effect grows faster than their positive knowledge spillovers.

Together, these results suggest that the details of firm heterogeneity make a first-order contribution to the optimal design of innovation policy in both sectors. I further compare the govern-

[8] Throughout the paper, I assume that the government can observe the productivity of the firm, and abstract away from potential information asymmetry between the government and the firm in this regard. In theory, we can infer firm productivity based on the observations of firm inputs and outputs. In practice, however, we need to account for the possibility that firms might distort their input and output decisions in response to the policy. Such considerations impose further information-related constraints on the policy, giving rise to second-best policies that account for moral hazard and screening issues (see, e.g., Akcigit et al., 2016). I leave the study of the second-best policies for future work.

[9] In comparing retail and manufacturing, I have been inspired by the salient differences in the patterns of productivity growth that prior work has documented between the two sectors. For instance, case studies in the US retail sector often attribute the transformative innovations that have fueled the fast growth of productivity in this sector to mature and productive incumbents, particularly big-box stores such as Walmart and Target (e.g., Foster et al., 2016; Basker, 2016). In this sector, firm productivity is highly correlated with measures of innovation such as IT investments (Doms et al., 2004). This pattern stands in contrast to manufacturing, where we typically associate transformative innovations with newcomers. In addition, firm productivity is highly correlated with measures of innovation such as IT intensity in retail, and reallocations play a relatively larger contribution to aggregate productivity growth in retail compared to manufacturing (Doms et al., 2004; Foster et al., 2006). In the model, retail's distinctive growth experience is compatible with an environment that features high innovation costs, high business volatility, and weak knowledge spillovers. These structural differences in turn reflect in the design of innovation policy, as we will see in the paper.

ment's spending on innovation subsidies under two different policies: a productivity-dependent policy that implements the optimal allocation, and the best among all one-size-fits-all, uniform subsidies that achieve the same rate of aggregate growth. I find that the productivity-dependent policy results in sizable savings in the spending of the government in both retail and manufacturing. Therefore, under this calibration, ignoring the details of firm heterogeneity appears to be fairly costly for the government.

Prior Literature This paper unites theories of firm heterogeneity and dynamics with the literature on firm innovation and growth. The former class of theories has highlighted the role of firm-level dispersion and volatility as important determinants of aggregate productivity (Jovanovic, 1982; Hopenhayn, 1992; Ericson and Pakes, 1995; Restuccia and Rogerson, 2008). This framework has found many applications in the empirical studies that link micro-level firm dynamics with macro-level outcomes.[10] However, it has primarily focused on understanding the *level* of aggregate productivity rather than its *rate of growth.* Instead, I follow the literature on endogenous growth theory in taking a stance on knowledge spillovers as drivers of long-run productivity growth. A number of recent papers have similarly studied growth within this framework, where competition across firms takes the form of monopolistic competition and spillovers work through knowledge diffusion, including Luttmer (2007, 2012), Lucas and Moll (2014), Perla and Tonetti (2014); Benhabib et al. (2014); Perla et al. (2015), and Sampson (2016). However, the current paper is the first to simultaneously endogenize and explain entry, exit, and the process innovation decisions of all incumbent firms, and to therefore provide a framework for policy analysis.[11]

As mentioned earlier, the core results of the paper contribute to a burgeoning literature that studies the normative implications of firm heterogeneity for the design of innovation policy. In particular, it complements the results of the recent paper by Atkeson and Burstein (2015), who emphasize the key role of the "social rate of depreciation" of innovation investments in the implications of any theory of growth for innovation policy. This rate captures our assumptions about the expected lifetime of ideas in the industry and, for instance, is zero in the workhorse model of innovation by Klette and Kortum (2004). Atkeson and Burstein show that the assumptions any model makes about the rate of social depreciation shapes the response of industry growth to the changes in the *aggregate* level of innovation investments. The current paper shows that the rate of social depreciation also shapes the response of growth to changes in the *distribution* of innovation investments across firms, if depreciation is due to the endogenous obsolescence of low-productivity ideas.

The displacement effect identified in the current paper is related to, but distinct from, the heterogeneity in externalities found in the model of Lentz and Mortensen (2014). They examine innovation policy in an extension of the Klette and Kortum (2004) theory with heterogeneous in-

[10]This formulation of firm innovation behavior goes back to the seminal work of Griliches (1979) and has been widely used in theoretical and empirical work in the fields of Industrial Organization and International Trade. Using this formulation, Olley and Pakes (1996) built an empirical model of industry evolution that provided a rigorous framework for the estimation of the dynamics of productivity across firms. A large literature has used this framework to link micro-level firm outcomes to the aggregate growth of productivity (e.g., Xu, 2008; Aw et al., 2011; Doraszelski and Jaumandreu, 2013; De Loecker, 2013).

[11]Knowledge diffusion and imitation feature in a number of other theories of industry dynamics, including, among others, Jovanovic and Rob (1989); Jovanovic and Macdonald (1994a), Acemoglu et al. (2006), Alvarez et al. (2008) Comin and Mestieri (2014), and König et al. (2015).

novative types for firms. In their theory, firms are able to innovate in as many product lines as they produce in. When low-innovative type firms innovate, they might randomly destroy the innovation option of a high-innovative type. This generates a negative externality for the former's innovation. However, this effect only applies to uninnovative (and typically unproductive) firms in their paper, whereas the displacement effect in the current paper can also grow with the productivity and size of the innovator.

Finally, the current paper focuses on process innovation and takes a firm's productivity as the measure of its stock of knowledge. This contrasts with the workhorse innovation model of Klette and Kortum (2004), in which all innovations are product innovations and the firm's number of products is the measure of its stock of knowledge. The distinction is perhaps best manifested in the results of the closely related paper by Acemoglu et al. (2013). They include endogenous exit and selection as an independent channel in an extension of the Klette and Kortum (2004) theory with heterogeneous innovative types.[12] However, the main interactions between selection and innovation that are the focus of this paper do not arise in their paper. Following the workhorse model, they also assume that all innovations are product innovations.[13] Therefore, the productivity of the innovator is independent of the expected lifetime of the ideas. In contrast, both types of externalities identified in the current paper stem from the link between firm productivity and the expected lifetime of the ideas it generates. As a result, the design and details of optimal innovation policy in the two papers are different.[14]

The remainder of the paper is organized as follows. In Section 2, I present the details of the model and characterize its long-run market equilibrium. In Section 3, I characterize the long-run optimal allocation of innovation and production in the model and present the results on the decomposition of externalities. In Section 4, I present the details and the results of the calibration exercise.

2 Model

2.1 Households and Demand

A mass $N_t = N_0 e^{g_N t}$ of households populate the economy and live in an infinite horizon. They choose the time paths of their consumption to maximize dynastic utility defined according to $U_o = \int_0^\infty e^{-(r-g_N)t} \log q_t \, dt$, where r and g_N stand for the discount rate of households and the rate of population growth, respectively. Per-capita consumption q_t of the household at time t is a CES

[12] As with the current paper, they also find that a lump-sum tax on the operation of firms combined with subsidies to innovation and entry can substantially raise growth. Quantitatively, the two papers find roughly similar extents of overall inefficiency in growth.

[13] This undirected account of innovation in the workhorse model also implies a thick-tailed distribution for the growth rates of firm employment. In a recent empirical investigation, Garcia-Macia et al. (2015) find that models with directed innovation, such that firms generate new know-hows that are similar to their existing ones, may better match the observed dynamics of firm employment in the US data.

[14] For completeness, I note that the growth model of Peters (2011) also generates endogenous market equilibrium misallocations across firms, as does the current paper. However, the misallocations in that paper stem from variations in markups. Here, the assumption of a CES demand structure implies that markups are constant across firms.

composite of consumption over a set Ω_t of products or services

$$q_t^{\frac{\rho}{1+\rho}} = \int_{\omega \in \Omega_t} q_t(\omega)^{\frac{\rho}{1+\rho}}\, d\omega,$$

where $\rho + 1$ is the elasticity of substitution. Normalizing household per-capita expenditure to unity at all times, we find the standard demand system $q_t(\omega) = P_t^{\rho}\, p_t(\omega)^{-(1+\rho)}$, where the price index P_t is given by

$$P_t^{-\rho} = \int_{\omega \in \Omega_t} p_t(\omega)^{-\rho}\, d\omega. \tag{1}$$

Households inelastically supply one unit flow of labor per person. Labor is homogeneous and earns wage rate w_t at time t. Households can also hold equity stakes in a balanced portfolio of all entrepreneurial activity and profit-making firms. Let a_t denote the per-capita assets and r_t be the interest rate at time t. The household problem involves the maximization of the dynastic utility subject to the flow budget constraint $\dot{a}_t = w_t - 1 + \mathrm{r}_t a_t - g_N a_t$ and a No-Ponzi condition. The Euler equation $\mathrm{r}_t = r$, household's initial asset endowment a_0, and the transversality condition $\lim_{t \to \infty} a_t e^{-(r-g_N)t} = 0$ together determine the time paths of household consumption.

2.2 Firms and Innovation

A continuum mass of producers of final goods or services form the industry. Each firm produces a single differentiated product ω and monopolistically competes against other firms active in the industry. Labor is the only factor of production and firms have to pay a flow cost of ψ_f (in units of labor) to continue operation. Firms are heterogeneous in terms of performance. I assume that all of this heterogeneity is captured by a single measure θ, which specifies the labor productivity of firms. Accordingly, the output of a firm θ is given by $q = \theta l$ where l is the number of its production workers.

Innovation Production Function Firm labor productivity θ_t is also a measure of the quality of its ideas at time t. I assume that the productivity stochastically evolves according to a stochastic process

$$d\theta_t = \Gamma_t\, dt + \sigma\, \theta_t\, d\mathcal{W}_t, \tag{2}$$

where $\mathcal{W}_t$ denotes a standard Wiener process and σ is the volatility parameter of the idiosyncratic productivity shocks, which arrive at the firm level. The drift of the process Γ_t is the outcome of the firm's instantaneous innovation investments, which raise its expected productivity in the future (Ericson and Pakes, 1995; Atkeson and Burstein, 2010). The innovation production function of the firm

$$\Gamma_t = G(i_t, \theta_t; l_t),$$

takes as inputs its innovation investments i_t (in units of labor) and its current ideas θ_t. In addition, it also depends on the firm's number of production workers l_t. Correspondingly, we can define an innovation cost function[15]

[15] Atkeson and Burstein (2010) assume the following slightly different specification that also delivers Gibrat's law. They assume $\Gamma_t = \theta_t^{\rho} \varphi\left(\frac{\Gamma_t}{\theta_t}\right)$. In contrast to their specification, the formulation of innovation costs here does *not* depend

$$i_t = \Phi(\Gamma_t, \theta_t; l_t) = l_t\, \varphi\left(\frac{\Gamma_t}{\theta_t}\right), \qquad \text{for } \; l_t > 0, \tag{3}$$

where $\varphi(\cdot)$ is a convex and monotonically increasing function, and satisfies $\varphi(0) = 0$ as well as a few other properties to be specified below. I assume that a firm that does not employ any workers cannot grow its productivity, i.e., $G(i_t, \theta_t; 0) = 0$.

Equation (3) assumes that innovation costs depend on the number of production workers. The idea is that in order to improve production processes in the firm, production workers at some level have to be involved in the implementation, experimentation, or training stages. Therefore, a firm that hires a larger number of employees has to pay a higher cost to modify its existing processes or adopt new technologies or managerial practices. Equation (3) also assumes that the current stock of firm ideas θ_t proportionally lowers the costs of improving productivity, which is a common assumption in theories of firm growth (e.g., Klette and Kortum, 2004). Together these two assumptions deliver Gibrat's law, which states that firm growth is independent of firm size and sales for the largest firms.

Production Decisions The simplest way to characterize the production decisions of firms is to consider a firm that has decided on a rate of productivity growth specified by $\gamma_t \equiv \Gamma_t/\theta_t$. Henceforth, I will refer to γ as the firm's innovation investment to simplify the exposition. Conditional on the innovation decision, we can specify the firm's employment, output, and pricing decisions. The firm maximizes its flow profits

$$\pi_t(\theta, \gamma) \equiv \max_q p_t(q) \cdot q - w_t \frac{q}{\theta}(1 + \varphi(\gamma)) - w_t \psi_f, \tag{4}$$

where I have substituted for firm production input $l \equiv q/\theta$. As we already saw, $p_t(q) \equiv P_t^{\frac{\rho}{1+\rho}} q^{-\frac{1}{1+\rho}}$. In choosing its scale of production, the firm takes the costs of innovation into account as part of its labor costs. Accordingly, the firm prices its products according to $p_t = \frac{1+\rho}{\rho} w_t (1 + \varphi(\gamma)) / \theta_t$.

Next, we can derive the relation between firm productivity and its market share. With the assumptions laid out above, it follows that the sales of a firm with productivity θ_t at time t are proportional to $s_t(\theta, \gamma) = N_t ((1+\rho) P_t / \rho w_t)^\rho \theta^\rho J(\gamma)$ where I have defined the function[16]

$$J(\gamma) \equiv (1 + \varphi(\gamma))^{-\rho}. \tag{6}$$

Function $J(\cdot)$ characterizes the effective costs of innovation for firms: investment in innovation raises effective production costs and therefore lowers instantaneous sales.[17]

on the elasticity of substitution parameter ρ, which is a demand-side parameter. Another important difference is that Equation (3) does not feature market size effects, since innovation costs scale with both firm employment and sales. By removing market size considerations, here I focus on isolating the heterogeneity in innovation externalities across firms of different productivities only due to the expected lifetime of ideas.

[16] I assume that there exists some $\overline{\gamma}$ such that $\lim_{\gamma \to \overline{\gamma}} \varphi(\gamma) = \infty$, and that the function j defined here is strictly concave, which requires that, for all $\gamma \in (0, \overline{\gamma})$

$$\varphi''(\gamma) > (1+\rho) \frac{(\varphi'(\gamma))^2}{1 + \varphi(\gamma)}. \tag{5}$$

[17] Equation (6) implies that stronger (static) product market competition, as proxied by the substitutability of firm

Let us now define the aggregate-level measures of the industry. Let M_t denote the industry measure defined in the space of productivity θ, and assume that along an equilibrium path a firm with productivity θ chooses innovation investments $\gamma_t(\theta)$ at time t. We can use the expression for the industry price index (1) to find the sales, profits, and employment of firms. Define the total mass of firms $\overline{M}_t \equiv \int_{\Omega_t} d\omega = \int dM_t(\theta)$, and the aggregate (average) productivity $\overline{\theta}_t$ through

$$\overline{\theta}_t^{\rho} \equiv \frac{1}{\overline{M}_t} \int \theta^{\rho} J(\gamma(\theta))\, dM_t(\theta). \tag{7}$$

Then, we can write firm sales as

$$s_t(\theta,\gamma) = \frac{N_t}{\overline{M}_t} \left(\frac{\theta}{\overline{\theta}_t}\right)^{\rho} J(\gamma),$$

and firm profits and wage bills are given by

$$\pi_t(\theta,\gamma) = \frac{1}{\rho+1} s_t(\theta,\gamma) - \psi_f w_t, \tag{8}$$

$$w_t l_t(\theta,\gamma)(1+\varphi(\gamma)) = \frac{\rho}{\rho+1} s_t(\theta,\gamma). \tag{9}$$

The total firm employment is the sum of firm production, innovation, and overhead workers given by $l(1+\varphi)+\psi_f$. Firm innovation intensity, defined commonly as the innovation spending to sales ratio, is given by $\varphi\rho/(1+\rho)$.

The formulation provided here considers firm innovation as *productivity-enhancing ideas embodied in the firm* (similar to the approach taken by Ericson and Pakes, 1995; Atkeson and Burstein, 2010). Conceptually, it shares core features with the organizational capital approach, which views firms as "storehouses of information" (Prescott and Visscher, 1980). Novel ideas improve a firm's knowledge of demand and technological environment, information on employee characteristics, managerial practices, organizational structure and culture, or any other competency that enables it to operate at a higher level of productivity.[18] This formulation contrasts with an alternative approach that considers innovation as embodied in the blueprints of new products developed by the firm (e.g., Klette and Kortum, 2004). When new ideas are embodied in the firm, as in the current model, the expected lifetime of the firm will matter for its incentives to invest in innovation. As I introduce the process of entry of firms in the next section, I will discuss another conceptual difference between the two approaches regarding the imperfection of transfer of knowledge across firms.

products ρ, raises innovation costs. But note that it also dynamically raises the future value of competition, as reflected in the marginal value of firm productivity.

[18] Syverson (2011) provides a broad classification of the factors that may influence firm productivity, including, firm-specific know-hows, management practices, and organizational structure. See Prescott and Visscher (1980) for examples of the first group, and Bloom and Reenen (2010) for a survey of the literature that establishes managerial practices as a key to the productivity and documents its heterogeneity across firms. A large body of work in the field of organizational economics has studied the impact of de jure processes and hierarchies, de facto norms, and firm leadership on the efficiency of a firm's operations (for recent reviews of the area, see the relevant chapters in the most recent handbook, e.g., Hermalin, 2012).

2.3 Entry

Entrepreneurs pay a flow ψ_e of costs (in units of labor) per unit flow of new firms. The initial productivity of a new firm is determined by the process of knowledge diffusion: an entrant builds on the knowledge available in the current pool of ideas in the industry (Luttmer, 2007, 2012; Sampson, 2016). I assume a general specification for this process that allows variations along two dimensions: (1) which incumbents an entrant is likely to adopt from, and (2) how much of the incumbent productivity can be transferred through adoption. Accordingly, once a new firm is born it receives the opportunity to adopt an idea from the pool of ideas currently available in the industry. The adoption distribution specifies the likelihood that this adopted idea may come from each of the firms in different tiers of the distribution of productivity. Once a firm adopts the idea of a given incumbent firm, the transfer of knowledge is characterized by the imperfection of adoption, as described below (for a similar account of diffusion, see Eeckhout and Jovanovic, 2002).

Adoption Distribution I assume that the likelihood of adopting the idea from any given incumbent firm depends on its current tier of productivity. We can think of $M_t/\overline{M}_t$ as a rank function mapping each tier of productivity to its rank in the current distribution of productivity. In particular, the exiting firm always has the zeroth rank $M_t(\theta_{o,t}) = 0$ and the rank converges to 1 as the productivity goes to infinity. I assume that, for entrants born at time t, the probability distribution of the adopted productivity θ_a is given by

$$\mathbb{P}(\theta_a \leq \theta) = F_a\left(\frac{M_t(\theta)}{\overline{M}_t}\right), \tag{10}$$

where F_a is a cumulative distribution function defined over the unit $[0, 1]$ interval. Distribution F_a accounts for the structural features of the industry that relate to the accessibility of different tiers of productivity for a potential entrant. In some industries, most entrants may be able to draw upon the ideas of firms close to the frontier, while in others they may be equally likely to draw upon lower-tier ideas.

As an example for the distribution of adoption, consider one in which the distribution is directed toward the more productive ideas. Let F_a be defined over $x \in [0, 1]$ as

$$F_a(x) = x^{\mu}, \qquad \mu \geq 1. \tag{11}$$

When $\mu > 1$, this distribution has a bias toward $x = 1$, since the mean of the distribution for the ranks is given by $\mu/(1+\mu)$. The corresponding probability density function f_a on the productivity ranks is monotonically increasing; we have $f_a(0) = 0$ and $f_a(1) = \mu$.[19]

If we assume $\mu = 1$, we find a uniform distribution F_a and we recover the common assumption

[19] We can think of distribution (11) as a special case of a beta distribution, which can produce a range of different means and dispersions over the unit interval. Here, I consider a subset of the parameter space by requiring that the likelihood of adoption from the frontier be nonzero, that is, $f_a(1) > 0$, which yields Equation (11) for the CDF of the distribution.

that entrants uniformly adopt ideas from the current distribution of productivities:

$$\mathbb{P}\left(\theta_a \leq \theta\right) = \frac{M_t\left(\theta\right)}{\overline{M}_t}. \tag{12}$$

For its simplicity, this particular form has been a popular choice in prior work (see, e.g., Perla et al., 2015; Atkeson and Burstein, 2015).[20]

Knowledge Transfer Once an entrant adopts the idea θ_a, the productivity θ_e with which it begins its production is given by

$$\theta_e = e^{\gamma_e}\,\widetilde{\theta}_t^{\eta}\,\theta_a^{1-\eta}, \tag{13}$$

where $\eta \in [0,1]$ is the degree of decreasing returns to adoption, γ_e is the innovation leap of entrants, and $\widetilde{\theta}_t$ is an industry-wide index of the quality of firm ideas. In what follows, I assume that $\widetilde{\theta}_t$ is the average productivity of the industry.[21]

When ideas are embodied in the structure and organization of firms, their transfer from one firm to another is likely to be imperfect. In this way, the specification above broadens the scope of innovation from those fully codifiable and transferrable, e.g., through patents, to ones that involve tacit and organizational knowledge.[22]

The formulation of entry in Equations (13) and (10) nests several previous models. If we assume $\eta = 0$, we find the case common in standard models of growth, in which knowledge perfectly transfers from incumbents to entrants (e.g., Luttmer, 2007; Perla et al., 2015). For instance, assuming an atomic adoption distribution with all the weight concentrated on the zeroth rank, we find the specification of Luttmer (2012). If we consider the case of $\eta = 1$ and further assume a stochastic distribution on γ_e, we find the formulation of entry of Sampson (2016) (when $\widetilde{\theta}_t$ is the average industry productivity), or the no-spillover formulation of Atkeson and Burstein (2010) (when $\widetilde{\theta}_t$ is a constant). The intermediate cases with $\eta \in (0,1)$ characterize situations where the entrants only imperfectly adopt the ideas of incumbents.[23]

2.4 Stationary Constant Growth Paths

In this section, I will characterize the asymptotic behavior of the industry when aggregate productivity grows at a constant rate and the distribution of firm size is stationary. Since the focus of

[20] The assumption of uniform entry does not match empirical facts on the small size of new entrants and their high likelihood of exit relative to the average firm in the industry (Dunne et al., 1988; Arkolakis, 2016), unless if we include decreasing returns to adoption as in Equation (13). Nevertheless, it provides a reasonable theoretical benchmark against which we can compare other possibilities.

[21] In the Appendix, I generalize this to an aggregator function that combines the current distribution of ideas with varying degrees of substitutability between the ideas of firms.

[22] This account of innovation closely matches with the concept of investments in intangible capital. Intangible investments are those costs that firms incur in product design, marketing, and organizational development (Corrado et al., 2009; Aizcorbe, 2009; Corrado and Hulten, 2010).

[23] Note that, as with the framework of Atkeson and Burstein (2015), the framework suggested here can also nest models where the long-run rate of growth of the economy is exogenous, and firm innovation only affects the long-run level of aggregate productivity. This is the case where we assume $\widetilde{\theta}_t$ is an exogenously growing function, which will then pin down the long-run rate of growth of aggregate productivity. The mechanisms discussed here qualitatively generalize to this case as well, but their quantitative size may be different.

the paper is on the asymptotics, I will defer many of the details to the appendix and only present the structure required for the definition of the stationary constant growth paths. Appendix 7.1 sets up the dynamic problem of an incumbent firm choosing its innovation γ and exit threshold, as well as a cohort of entrants choosing their rate of entry. Combining the two, it defines the industry equilibrium path and the general equilibrium of the economy.

The next definition formally introduces the concept of a stationary constant growth path. We can think of this as the heterogenous-firm counterpart to the steady state of standard growth models. Among the potential equilibrium paths, it limits attention to the set that is of empirical interest.[24]

Definition 1. *(Stationary Constant Growth Path Equilibrium)* An equilibrium path along which a nonzero flow of entrants continuously arrive in the industry, and,

1. The mass and average productivity of firms asymptotically grow at constant rates, that is, $\lim_{t\to\infty} \dot{\overline{M}}_t / \overline{M}_t = g_M$ and $\lim_{t\to\infty} \dot{\overline{\theta}}_t / \overline{\theta}_t = g_\theta$,
2. The distribution of firm size converges to a stationary distribution $\lim_{t\to\infty} \mathbb{P}\left\{l_t\left(\theta\right) \le l\right\} = \mathbb{P}_l\left(l\right)$ with finite mean.

To characterize the stationary constant growth path, I need to transform the productivity states into a space where their distribution remains stationary. Since by definition the distribution of firm employment is stationary along such an equilibrium path, this transformed space has to be a one-one function of firm employment. Define the firm *efficiency* state z_t such that is satisfies[25]

$$e^{z_t(\theta)} \equiv \frac{1}{\left(\rho+1\right)\psi_f w_t} \frac{N_t}{\overline{M}_t} \left(\frac{\theta}{\overline{\theta}_t}\right)^{\rho} \tag{14}$$

This state is the ratio of the gross profits to fixed costs for a θ-productivity firm that chooses *not* to invest in productivity growth (hence, $J = 1$). It captures the position of the firm relative to the other firms in the industry. A firm with efficiency $z = 0$ earns gross profits that exactly compensate for its fixed costs.

Along a constant growth path, the wage rate converges to a constant, since per-capita expenditure is normalized to unity. We can show that a necessary condition for the existence of a stationary constant growth path is that the rate of growth of the total mass of firms is the same as the rate of growth of population $g_M = g_N$ (see Appendix 7.1). Since the average firm employment remains constant and since firm employment and profits are proportional in a stationary distribution, the growth in aggregate sales has to be absorbed by the rising mass of firms.

[24]Other potential paths are those that feature, e.g., no entry (Luttmer, 2012, characterizes such equilibria in a setting similar to the one here), or nonstationary distributions of firm size with explosive dispersion. Empirical work has typically found stable distributions of firm size across a wide range of industries (Luttmer, 2010), and broad industries such as those that are the focus of this paper almost always exhibit continuous entry of new firms.

[25]It is straightforward to show that under the condition $\lim_{t\to\infty} \dot{\overline{M}}_t / \overline{M}_t = g_N$, the distribution of any monotonic function of $\theta / \overline{\theta}_t$ will be stationary along a stationary constant growth path, and therefore all such functions provide alternative transformations. The particular choice in definition (14) is similar to the one used by Luttmer (2012) and helps simplify the form of the differential equation characterizing the value of function of firms along a stationary constant growth path.

Since $g_M = g_N$ and the wage rate w_t is asymptotically constant, the evolution of firm efficiency states along a stationary constant growth path is given by the brownian motion process

$$dz_t = \rho \left(\gamma - g_\theta - \frac{1}{2}\sigma^2 \right) dt + \rho\,\sigma\, d\mathcal{W}_t. \tag{15}$$

This expression shows that we can define the *rate of profit erosion* $g \equiv \rho\left(g_\theta + \frac{1}{2}\sigma^2\right)$ as the rate at which growth in the aggregate productivity erodes the profits of any firm active in the industry. Alternatively, we can think of this as the rate of depreciation of ideas in the industry. As we might expect, this rate depends on the rate of aggregate productivity growth g_θ, volatility σ, and the index of market competition ρ.

Along a stationary constant growth path, we can write the evolution of the industry measure as

$$M_t(\theta) = \overline{M}\, e^{g_M t} \times H\left(z_t(\theta)\right), \tag{16}$$

where H is a cumulative distribution function. The measure asymptotically exhibits the so-called *traveling-wave* behavior, since it is a stationary function of $\log\theta - g_\theta t$ (Lucas and Moll, 2014). The total mass of firms grows at a constant rate, and their productivity distribution travels at a constant rate g_θ in the log-productivity space. The last condition in the definition of the stationary constant growth path requires that the moment $\mathbb{E}_H\left[e^z J\right]$ should be finite, ensuring that a well-defined average size of firm employment exists (this is the only empirically relevant case, see Luttmer, 2010).

The remainder of this section studies the behavior of firms and the aggregate industry along a stationary constant growth path. I will first characterize the firm value function and optimal policies, and then characterize the distribution of firm productivity H. Finally, I combine the two and examine how the rate of profit erosion g is determined through the zero entry condition.

2.5 Incumbent Problem

Along a stationary constant growth path, we can write the value of a firm with efficiency state z as $\psi_f\, w_t V(z)$. So long as the firm continues to operate, standard arguments imply that the value function $V(z)$ satisfies the stationary Hamilton-Jacobi-Bellman (HJB) equation

$$rV(z) = \max_{\gamma}\, e^z J(\gamma) - 1 + \rho\left(\gamma - g_\theta - \frac{1}{2}\sigma^2\right) V'(z) + \frac{\rho^2}{2}\sigma^2\, V''(z), \tag{17}$$

for $z \geq z_o$ where z_o is the exit threshold and the function J is defined by Equation (6). The following proposition characterizes the solution to this problem. The derivation of the stationary HJB equation and the proof of the proposition in a more general setting are included in Appendix 7.1.

Proposition 1. *Assume that the innovation cost function $\varphi(\cdot)$ satisfies the conditions stated in Equation (16), and that the rate of productivity growth g_θ is large enough so that*

$$r + \rho\, g_\theta - \frac{\rho}{2}(\rho - 1)\,\sigma^2 > \rho\overline{\gamma} - \frac{J(\overline{\gamma})}{J'(\overline{\gamma})}. \tag{18}$$

Then the HJB equation (17) has a unique, continuous, convex, and monotonically increasing value function $V(\cdot)$ that converges to a stable, asymptotic function $\lim_{z\to\infty} V(z) = u^* e^z$. Moreover, the corresponding policy function $\gamma^*(z)$ is unique, continuous, and monotonically increasing. It satisfies $\gamma^*(z_o) = 0$ for exiting firms, and an asymptotic Gibrat's law, in the sense of $\lim_{z\to\infty} \gamma^*(z) = \gamma^{**}$, for the largest firms. The asymptotic rate of productivity growth of the largest firms γ^{**} satisfies the maximum in the following problem

$$u^* = \max_{\gamma} \quad \frac{(1+\varphi(\gamma))^{-\rho}}{r+\rho\left(g_\theta - \gamma - \frac{1}{2}(\rho-1)\sigma^2\right)}. \tag{19}$$

Proof. See Appendix 7.1. □

Proposition 1 establishes the conditions that ensure the existence of a uniquely stable equilibrium for the growth strategies of incumbent firms. The key requirement involves a lower bound on the effective discount rate of firms. This result is intuitive: since productivity gains are permanent, the discount rate has to be large enough to contain firms from infinite investments in growth.

The convexity of the value function follows from the fact that firms face different expected lifetimes of activity in the industry. Productive firms lie farther from the exit cutoff and expect to live longer in the industry. A marginal increase in the productivity of a firm raises its value along two margins: it increases the flow of profits as well as the expected lifetime of the firm. This implies that the marginal net present value of productivity growth is larger for larger firms. Due to their shorter expected lifetimes, unproductive firms have less to gain from investing to permanently raise their efficiency. Facing higher returns to investment, large, high-efficiency firms invest more than small, low-efficiency firms, giving rise to an increasing investment policy function $\gamma^*(\cdot)$. The proposition further shows that the value function becomes linear in e^z in the limit of very large firms. As a result, the positive relationship between productivity (size) and innovation investments declines for very large firms and we find that Gibrat's law holds in the limit.

As expected, factors that contribute to higher discounting of the net present value of incumbent profits have a negative impact on the firms' value and innovation investments. In particular, the rate of aggregate productivity growth g_θ negatively affects the firms' incentives to grow.[26] To see this, consider a firm that is forced to remain active at all productivity levels. The value function of this firm is simply given by $J(\gamma^{**}) e^z/r^* - 1/r$ where γ^{**} is the firm's productivity growth and $r^* = r + \rho\left(g_\theta - \gamma^{**} - \frac{1}{2}(\rho-1)\sigma^2\right)$ is its effective discount rate. Increasing r^*, say, through an increase in the rate of productivity growth g_θ, erodes the future value and therefore the rate of return to innovation investments. For high-productivity firms that have large efficiency z, the option to exit the industry is not relatively valuable. Therefore, this analysis provides a good approximation for the effect of faster growth on their incentives to innovate.

For less productive firms that lie close to the exit cutoff, faster growth creates additional considerations since it affects their valuation of the option to exit. Of course, the option to exit the

[26] The same negative relationship between the rate of creative destruction and the incentives to grow exists in canonical models of innovation and growth that feature both entry and incumbent innovation (see, e.g., Klette and Kortum, 2004; Acemoglu and Cao, 2015). The mechanism in both cases is the erosion of future profits along the growth path, which lowers the expected private lifetime of benefiting from innovation. The main difference is that here the effect is stronger on low-productivity firms, whereas it is uniform under the canonical model.

industry is relatively more valuable for these firms. A rise in the rate of productivity growth g_θ changes the option value of exit since it affects first, how fast the relative position of a firm in the industry erodes, and second, at which efficiency level it ceases operation (the exit cutoff). The latter effect only emerges when the extent of volatility is high and low-productivity firms account for the option value of continuation in their exit decisions. Together, these forces imply that the dynamic displacement effect of faster growth creates heterogeneous responses in the innovation investments of firms.

2.6 Industry Problem

As we saw in Section 2.7, the mass of active firms at time t along a stationary constant growth path takes the form $M_t(z) = \overline{M} e^{g_M t} H(z)$, where $H(\cdot)$ is the cumulative distribution of firm efficiencies. This stationary distribution emerges as the result of the innovation and exit decisions of all firms, and characterizes the long-run state of the industry. In this section, I characterize this distribution and provide conditions required for its existence.

Defining the industry-wide index of firm efficiency through

$$e^{\tilde{z}} \equiv \left(\int e^{z/\rho} h(z)\, dz \right)^{\rho},$$

we can transform the knowledge transfer Equation (13) to the space of firm efficiencies as

$$z_e = \eta\, \tilde{z} + (1-\eta)\, z_a + \rho \gamma_e.$$

The distribution of the efficiency with which entrants attempt entry is then given by $F_e(z) \equiv H\left(\frac{z - \eta \tilde{z} - \rho\gamma_e}{1-\eta}\right)$.

Along a stationary equilibrium path, the exit cutoff for firms corresponds to a cutoff in the space of efficiency. Let z_o denote this cutoff, and $\underline{z}_a$ denote the lowest level of incumbent efficiency that an entrant can adopt to successfully enter the industry, satisfying $\eta\tilde{z} + (1-\eta)\, \underline{z}_a + \rho\gamma_e = z_o$. If $\rho\gamma_e < -\eta\,(\tilde{z} - z_o)$, this level of efficiency corresponds to the one of an active incumbent firm; therefore, a flow of firms do not succeed in their attempt to enter the industry and immediately exit. In this case, the gross rate of entry is larger than the rate of successful entry by a factor of $[1 - H(\underline{z}_a)]^{-1}$. Otherwise, all entry attempts are successful. We can accordingly define a distribution of entry efficiency F with a cumulative distribution function

$$F(z) = \begin{cases} \frac{F_e(z) - F_e(z_o)}{1 - F_e(z_o)}, & \rho\gamma_e < -\eta\,(\tilde{z} - z_o)\,, \\ F_e(z)\,, & \text{otherwise.} \end{cases}$$

Let $h(z) \equiv H'(z)$ denote the probability density function for the stationary distribution, and λ_e and λ_o be the rates of (successful) entry and exit. The distribution H then has to satisfy the following stationary Kolmogorov Forward Equation (KFE) for $z > z_o$

$$g_M H(z) = \lambda_e F(z) - \lambda_o - \rho \left(\gamma^*(z) - g_\theta - \frac{1}{2}\sigma^2 \right) h(z) + \frac{\rho^2}{2}\sigma^2 h'(z)\,, \tag{20}$$

along with boundary conditions $h(z_o) = 0$ and $\lim_{z\to\infty} H(z) = 1$, where $\gamma^*(\cdot)$ is the incumbent investment policy solving problem (17). As z goes to infinity (zero), the cumulative and probability distribution functions H and h converge to 0 and 1 (0 and 0), respectively. Equation (20) then implies a rate of exit given by

$$\lambda_o = \frac{1}{2}\rho^2\sigma^2 h'_+\left(z_o\right), \tag{21}$$

and a rate of (successful) entry λ_e given by the conservation of mass condition $\lambda_e = \lambda_o + g_M$.

The empirically relevant equilibria are those in which the industry features both a stable and stationary distribution of firm size and a constant stream of entrants and exiting firms. Therefore, it is natural to search for industry equilibria that feature this property. The following proposition gives us the necessary and sufficient conditions required to establish the existence of a stationary distribution of firm efficiency and size along with a positive rate of entry. It also establishes that the distribution will always have a right tail that exhibits Pareto behavior.

Proposition 2. *(Stationary Distribution) Assume that at least one of the two following conditions holds: either adoption is imperfect $\eta > 0$, or the likelihood of adoption from the frontier is zero $f_a(1) = 0$.*[27] *Then, for a stationary distribution of firm size with a positive rate of entry λ_e and finite mean to exist, the rate of growth in the mass of firms should be large enough, in the sense that $g_M > 0$ and $g_M > \rho\left(\gamma^{**} - g_\theta\right) + \frac{1}{2}\rho\left(\rho - 1\right)\sigma^2$. The rate of entry for any other potential stationary distribution will necessarily be zero (if such a distribution exists). The right tail of the distribution exhibits a Pareto behavior with a tail index given by*[28]

$$\zeta \equiv \lim_{z\to\infty} \frac{h\left(z\right)}{1 - H\left(z\right)} = \frac{1}{\rho}\left[\sqrt{\left(\frac{g_\theta - \gamma^{**}}{\sigma^2} + 1\right)^2 + \frac{2g_M}{\sigma^2}} + \frac{g_\theta - \gamma^{**}}{\sigma^2} + 1\right]. \tag{22}$$

Proof. See Appendix 7.1. □

Proposition 2 imposes a lower bound on g_M in order to establish the existence of the stationary distribution. The intuition is also similar to the condition we found in Proposition 1: we require a flow of entrants that is large enough to prevent the most productive firms from indefinitely growing their share of employment (Luttmer, 2012). Furthermore, it requires that no entrants receive spillovers directly from the frontier technology, or if they do, then the .

2.7 Equilibrium Industry Growth

Combining Propositions 1 and 2, we can now close the model through a free entry condition that pins down the rate of productivity growth in the long-run equilibrium.

I assume that the rate of growth of population (aggregate sales) is positive $g_N > 0$. As discussed earlier, the mass of firms grows at the rate $g_M = g_N$ and therefore the first condition for the existence of the stationary distribution in Proposition 2 is satisfied. Defining the relative entry to

[27] In the case where adoption is perfect $\eta = 0$ and $f_a(1) > 0$, a necessary condition for the existence of a stationary equilibrium is that $\gamma_e < 0$. In addition, the rate of growth in the mass of firms g_M has to satisfy a more stringent condition to ensure the existence of an equilibrium. The characterization of this case is available upon request.

[28] Note that $-\zeta$ is the negative root of the quadratic characteristic equation $\mathsf{Q}(x) = \frac{1}{2}\rho^2\sigma^2x^2 + \rho\left(\gamma^{**} - g\right)x - g_M$, corresponding to the KFE characterizing the stationary distribution.

fixed costs $\psi \equiv \psi_e/\psi_f$, we can write the free entry condition as

$$\psi = \int V\left(\eta\tilde{z} + (1-\eta)z + \rho\gamma_e\right) f\left(H(z)\right) h(z)\, dz = \left(1 - F_e(z_o)\right) \mathbb{E}_F\left[V(z)\right], \tag{23}$$

where the pre- and post-entry distribution of efficiencies F_e and F are defined as in the previous section. Equation (23) pins down the rate of productivity growth g_θ, which is ceteris paribus decreasing in the ratio of entry to fixed costs ψ.

All the conditions required for the existence of the value function and the stationary distribution together translate into a constraint on the relative costs of entry, in the form of $\psi \in \left(\underline{\psi}, \overline{\psi}\right)$. Such a condition ensures, for any relative cost ψ within the interval, the existence of a unique stationary distribution H of efficiencies with nonzero rate entry λ_e, such that

$$\lambda_e = g_N + \frac{1}{2}\rho^2\sigma^2 h'_+(z_o).$$

The relative market share of a firm in state z in the aggregate sales of the industry is given by

$$\nu(z) = \frac{e^z J\left(\gamma^*(z)\right)}{Z}, \tag{24}$$

where I have defined an index of average firm size $Z \equiv \mathbb{E}_H\left[e^z J\left(\gamma^*(z)\right)\right]$. Variable ν compares the sales of a firm at efficiency level z with the average sales of the industry; therefore, it averages to one across the industry , $\mathbb{E}_H\left[\nu(z)\right] = 1$. Therefore, market shares $\nu(\cdot)$ define a distribution, which I also denote by ν_s. The expectation of this distribution over any firm-level outcome $X(\cdot)$ in the efficiency space defines a market-share-weighted average of that outcome given by $\mathbb{E}_\nu\left[X(z)\right] \equiv \mathbb{E}_H\left[e^z J\left(\gamma^*(z)\right) X(z)\right]/Z$.

The following proposition provides a decomposition of the rate of average productivity (and hence the rate of profit erosion) to the components stemming from incumbent investment as well as the margin of entry and exit.

Proposition 3. *(Growth Decomposition) The overall rate of profit erosion of the industry can be decomposed into*

$$\begin{aligned} \mathbb{E}_\nu\left[1 + \varepsilon_{J,1}\right] \cdot g &= \rho\, \mathbb{E}_\nu\left[\gamma^*(z)\left(1 + \varepsilon_{J,1}\right)\right] + \frac{\rho^2}{2}\sigma^2\left(\mathbb{E}_\nu\left[\left(1 + \varepsilon_{J,1}\right)^2\right] + \mathbb{E}_\nu\left[\varepsilon_{J,2}\right]\right) \\ &\quad + \left(\nu_e - \nu_o\right)\cdot\lambda_o - \left(1 - \nu_e\right)\cdot g_M \end{aligned} \tag{25}$$

where ν_e and ν_o denote the market shares of entrants and exiting firms, given by $\mathbb{E}_F\left[e^z J\right]/Z$ and e^{z_o}/Z, respectively, and ε^J_1 and ε^J_2 denote first and second order semi-elasticities of the innovation costs with respect to efficiency, given by $\varepsilon_{J,1} \equiv \frac{\partial}{\partial z}\log J\left(\gamma^*(z)\right)$ and $\varepsilon_{J,2} \equiv \frac{\partial^2}{\partial z^2}\log J\left(\gamma^*(z)\right)$.

Proof. See Appendix 7.1. □

The decomposition in Equation (25) has a straightforward interpretation. The first term on the right hand side gives the share-weighted growth rates of firms at different efficiency states. This term is driven by the endogenous investments of incumbent firms. The second term captures the contribution of volatility to growth. Idiosyncratic shocks may also contribute to aggregate growth as market shares reallocate toward firms that receive better shocks. We can interpret this term as the contribution of involuntary experimentation of firms in the industry. Together, the two expressions on the first line summarizes the contribution of incumbent firms to productivity growth.

The third and fourth terms, on the second line of the equation, summarize the contribution of entry and exit. The third term captures the contribution of selection (entry and exit). New firms replace firms at the exit threshold at a rate given by the exit (turnover) rate λ_o. Each new entrant improves the efficiency of an exiting firm by a factor $\left(\mathbb{E}_F\left[e^z J\right] - e^{z_o}\right)/Z$, that is, the difference between the average size of entrants and exiting firms, relative to the average firm size. This expression in turn depends on the gap in the productivity of the entrants and exiting firms.

Finally, the fourth term accounts for another distinct contribution of entry, stemming from a potential gap between the productivity of entrants and the average incumbent. The net rate of entry of firms is given by the rate of growth in the mass of firms g_M. Each new firm faces a deficiency of size $1 - \mathbb{E}_F\left[e^z\right]/Z$ relative to an average incumbent. Therefore, a faster rate of growth in the number of firms can reduce the overall rate of growth of productivity in the industry.

To more clearly see the intuition behind the decomposition, let us simplify it by assuming that $\gamma^*(z)$ is approximately constant across two high- and low-innovation regions of the efficiency space. Therefore, the terms involving the elasticities of the firm investment policies drop out and we find

$$g_\theta \approx \underbrace{\mathbb{E}_\nu\left[\gamma^*(z)\right] + \frac{1}{2}(\rho-1)\sigma^2}_{\text{Incumbents}} + \underbrace{\left(\frac{\nu_e - \nu_o}{\rho}\right)\cdot\lambda_o}_{\text{Turnover}} - \underbrace{\left(\frac{1-\nu_e}{\rho}\right)\cdot g_M}_{\text{Net Entry}}, \tag{26}$$

where $\mathbb{E}_\nu\left[\gamma^*(z)\right]$ is the market-share-weighted average of incumbent innovations. Innovation investments of incumbent firms influence the rate of productivity growth through two distinct channels: 1) directly, through the first term in Equation (26), and 2) indirectly, through their effects on the contribution of turnover. The second effect happens both because incumbent innovation raises the rate of turnover, and because it influences the gap between entering and exiting firms. Despite the fact that firms invest in improving their own processes, they still destroy the business of unproductive firms and push them out of the industry. Here, creative destruction operates through the selection margin.

Typically, any shock to the environment or any policy intervention shifts the composition of growth in the industry between the different components above. Consider again a small rise in the ratio of entry to fixed costs ψ. From the free entry condition, we expect the aggregate rate of profit erosion g to respond negatively to this shock. Equation (25) allows us to further quantify the resulting shift in the composition of growth. Based on what we saw from the behavior of incumbent investments, they will rise in response to this fall in the rate of profit erosion g. We also expect the rate of entry $\lambda_o + g_M$ to fall as a result of the rise in relative entry costs. Therefore, the role of incumbent investment in overall industry growth rises in response to this shock.

2.8 Solving for the Model

As with most models of investment under uncertainty, we cannot analytically characterize the solution to either the incumbent or the industry problem.[29] Instead, I employ the following scheme to compute numerical solutions to both problems.

In order to solve the Bellman equation in Proposition (1), I first transform the problem into a first-order system of differential equations $\boldsymbol{u}'(z) = \mathcal{Q}(\boldsymbol{u}(z))$ defined in a two-dimensional state space $\boldsymbol{u} \in \mathbb{R}^2_+$ where

$$u_1(z) = \left(V(z) + \frac{1}{r}\right) e^{-z}, \tag{27}$$

$$u_2(z) = V'(z)\, e^{-z}. \tag{28}$$

We have a Boundary Value Problem (BVP) with the following two boundary conditions: 1) the smooth pasting condition implies $u_2(z_o) = 0$, and 2) Proposition 1 implies $\lim_{z\to\infty} u_1(z) = u^*$ where u^* is given by equation (19). Figure 1 gives an example of how the solution typically behaves in the $\boldsymbol{u}$-space. The standard smooth-pasting condition implies that the path has to begin somewhere on the u_1-axis $\boldsymbol{u}(z_o) = (u_1(z_o), 0)^t$, which is located to the southwest of the equilibrium point such that $u_1' \leq 0$ and $u_2^{'} \geq 0$. From the standard value-matching condition $V(z_o) = 0$ and Equation (27), this point pins down the exit threshold

$$e^{z_o} = \frac{1}{r\, u_1(z_o)}.$$

For the infinite boundary, the solution has to converge to the point $\boldsymbol{u}^* = (u^*, u^*)$, which is an *equilibrium point* for the system of differential equations in the sense that $\boldsymbol{Q}(\boldsymbol{u}^*) = 0$. Condition (18) both ensures that such an equilibrium point exists, and that it is locally saddle path stable. In other words, the BVP has a unique solution, characterized by a path of $\boldsymbol{u}(z)$ that converges to $(u^*, u^*)^t$. Figure 1 shows both the stable and unstable paths around the equilibrium point.

I derive a similar scheme to solve for the stationary distribution of efficiencies $H(\cdot)$ for any given specification of the model. I transform the problem into a first-order BVP $\boldsymbol{h}'(z) = \boldsymbol{\mathcal{P}}(\boldsymbol{h}(z))$ defined in a two-dimensional state space of $\boldsymbol{h}(z) \equiv (H(z), h(z)) \in \mathbb{R}^2_+$. The problem has the following boundary conditions: $h(z_o) = H(z_o) = 0$ and $\lim_{z\to\infty} H(z) = 1$. Figure 2 gives an example of how the solution typically behaves in the $\boldsymbol{h}$-space. In the limit of large z, the path of the solution has to converge to the point $\boldsymbol{h}^* = (1, 0)^t$ that forms an equilibrium point for the system of differential equations in the sense that $\boldsymbol{\mathcal{P}}(\boldsymbol{h}^*) = 0$ (another equilibrium point $(0, 0)^t$ is also on the path). The conditions in Proposition 2 ensure that this equilibrium point is locally saddle path stable. Figure 2 further shows both the stable and unstable paths around the equilibrium point $\boldsymbol{h}^* = (1, 0)^t$.

[29] Nevertheless, we can use approximations to transform the two problems from differential equations into algebraic equations. In the online appendix I showcase different properties of the incumbent and the industry problems under a specific innovation cost function and an atomic adoption distribution through both numerical solutions and analytical approximations. However, even in this case, or eve cases where the value function can be analytically computed, the resulting expressions involve nonlinearities that do not give rise to transparent comparative statics. More generally, the difficulty in analytical study of comparative statics is a well-known feature of models of investment under uncertainty (Dixit and Pindyck, 1994). Dixit (1991) suggests analytical approximations based on Taylor expansions as an alternative approach.

I use different generalizations of the method above to solve for the variations of the Bellman equation encountered in the cases involving distortions and the social planner's problems.

3 Optimal Allocations

Let us now examine the allocative efficiency of the market equilibrium by comparing it against the allocations advised by a social planner. As mentioned in the Introduction, we find that the potential gaps between the social and private returns to innovation are substantially different in this model from those in standard models that do not include selection and knowledge diffusion. In particular, the externalities created by firm-level innovations on the growth of industry-level productivity vary with firm productivity.

Section 3.1 begins by characterizing the socially optimal allocation of innovation and production. Section 3.2 discusses the different channels through which innovation investments generate externalities on industry-level productivity growth in this model. Section 3.3 then briefly discusses the numerical scheme used for solving for the optimal allocation. Finally, Section 3.4 presents two different schemes that the social planner may employ to implement the optimal allocation, under the assumption that she can observe the productivity ranking of firms. The implementation allows us to study the sizes of gaps between social and private returns to innovation investments.

3.1 Socially Optimal Allocations

In the environment of this model, for incumbent firms at any given level of efficiency the social planner has to decide on 1) the extensive margin of production (selection), i.e., whether the firms should be selected to continue operation, 2) the intensive margin of production, i.e., how much output and labor input to allocate to them, and 3) innovation investments. In addition, the social planner decides on the rate of entry into the industry.

Conditional on the extensive margin of production and the choices of firm innovation investments, the first-order conditions corresponding to the static allocations of output and employment coincide between the market equilibrium and the socially optimal allocations. This result directly follows from the assumption of CES demand and constant markup pricing. Therefore, the intensive margin of production decisions (output and labor input) is efficient, conditional on the extensive margin (exit) and innovation investments. However, the extensive margin, whether the firm continues to operate or exits, may vary between the market equilibrium and the socially optimal allocation.[30] In addition, differences between the social and market allocations may arise in the innovation investment decisions of firms. I will denote the socially optimal efficiency cutoff for production as $z_{o,s}$, and the socially optimal innovation investments of firms as $\gamma_s(\cdot)$ for $z \geq z_{o,s}$.

[30] In the static or symmetric setting, it is well known that with CES preferences business stealing and appropriability effects exactly cancel out and the market entry (and exit) of firms and products becomes optimal (Grossman and Helpman, 1991, p. 82). The framework offered here is flexible enough to allow for distortions due to potential variations in markups that exist in more general demand specifications (e.g., Dhingra and Morrow, 2012).

The following proposition presents the key result that characterizes the socially optimal allocations.[31]

Proposition 4. *(Socially Optimal Allocations) Assume that the socially optimal allocation has a stationary constant growth path characterized by a rate of productivity growth $g_{\theta,s}$, the social value function $V_s(\cdot)$, the CDF and pdf of the stationary distribution of efficiencies $H_s(\cdot)$ and $h_s(\cdot)$, respectively, and the rate of entry $\lambda_{e,s}$. Social value function V_s satisfies a social HJB equation*

$$\begin{aligned} rV_s(z) &= \max_{\gamma} e^z J(\gamma) - 1 + \rho \left(\gamma - g_{\theta,s} - \frac{1}{2}\sigma^2 \right) V_s'(z) + \frac{\rho^2}{2}\sigma^2 V_s''(z) \\ &\quad + \lambda_{e,s} \cdot (U_s(z) - \Psi_s), \end{aligned} \tag{29}$$

for $z \in (z_{o,s}, \infty)$ and the boundary conditions $V_s(z_{o,s}) = V_s'(z_{o,s}) = 0$, where the social crowding cost Ψ_s and the spillover function $U_s(\cdot)$ are defined as follows. The social crowding and spillover are each composed of two separate contributions due to knowledge diffusion and to the industry-wide average stock of knowledge given by

$$\begin{aligned} U_s(z) &= \eta \left(e^{\frac{z - z_o}{\rho}} - 1 \right) \tilde{u}_s + (1 - \eta)\, u_s(z), \\ \Psi_s &= \eta \left(e^{\frac{\tilde{z} - z_o}{\rho}} - 1 \right) \tilde{u}_s + (1 - \eta)\, \psi_s, \end{aligned}$$

where the knowledge diffusion terms are given by

$$\psi_s \equiv \int_{z_{o,s}}^{\infty} V_s'(\eta \tilde{z} + (1 - \eta) x + \rho \gamma_e)\, f_a(H_s(x))\, (1 - H_s(x))\, dx, \tag{30}$$

$$u_s(z) \equiv \int_{z_{o,s}}^{z} V_s'(\eta \tilde{z} + (1 - \eta) x + \rho \gamma_e)\, f_a(H_s(x))\, dx, \tag{31}$$

and the term corresponding to the contribution to the industry-wide knowledge stock is given by

$$\tilde{u}_s \equiv \rho e^{-\frac{\tilde{z} - z_o}{\rho}} \int_{z_{o,s}}^{\infty} V_s'(\eta \tilde{z} + (1 - \eta) x + \rho \gamma_e)\, f_a(H_s(x))\, h_s(x)\, dx. \tag{32}$$

In addition, H_s satisfies a corresponding KFE and, if $\lambda_{e,s} > 0$, the socially optimal value function and efficiency distribution satisfy a social free entry condition

$$\psi = \int_{z_{o,s}}^{\infty} V_s(\eta \tilde{z} + (1 - \eta) z + \rho \gamma_e)\, f_a(H_s(z))\, h_s(z)\, dz, \tag{33}$$

where, as before, we have $\psi \equiv \psi_e / \psi_f$.

Proof. See Appendix 7.2. □

When the long-run equilibrium involves entry, the proposition shows that the distinction between the social and private value of firms appears in a constant Ψ_s, and an efficiency-dependent knowledge spillover term $U_s(\cdot)$ in the social planner's HJB equation. The terms corresponding to the stock of knowledge are standard spillover terms: the external contribution of firms to the

[31] The proposition provides only the necessary conditions for a constant growth path to characterize a socially optimal allocation in the long-run.

industry-wide knowledge stock is proportional to their productivity (that is, $e^{z/\rho}$). Raising this stock then generates social value by increasing the efficiency of all entrants, creating social value given by Equation (32).

For the remainder of this section, I focus on understanding expressions (30) and (31), corresponding to the contribution of knowledge diffusion. The two expressions capture simple intuitions about the workings of knowledge diffusion in this model. To better see these intuitions, let us consider a perturbation to the socially optimal (or the market equilibrium) allocation whereby we add a small mass Δm of firms with efficiency z to the long-run stationary distribution, removing a mass of the same size at random from the distribution. This small perturbation keeps the total mass of firms intact, but influences the ranking of firms and, accordingly, modifies the distribution of entrants. We can derive the social value of the knowledge spillovers created by this perturbation, that is, expressions (30) and (31), in three steps.

Step (I): Improving or Crowding of the Pool of Ideas First, consider an entrepreneur E who has received the draw of $\mu_E \in (0,1)$ as the ranking of the idea she adopts. The corresponding efficiency z_a then satisfies $\mu_E = H(z_a)$. The effect of the perturbation on her efficiency z_a depends on whether the efficiency z of the added firms is better or worse than her pre-perturbation adopted idea z_a. If the new idea is better $z > z_a$, this will improve the pool of ideas for entrepreneur E. This will change the rankings of ideas from the pre-perturbation $H(z_a)$ by a small amount

$$\Delta H \approx H(z_a) \times \frac{1}{1+\Delta m} - H(z_a) \approx -H(z_a)\,\Delta m.$$

Since her draw μ_E should remain constant, the perturbation pushes up her adopted efficiency by $\Delta z_a = H(z_a)/h(z_a)\,\Delta m$. On the other hand, if the new firms z are worse than her pre-perturbation idea $z < z_a$, this will deteriorate the pre-perturbation ranking by

$$\Delta H \approx H(z_a) \times \frac{1}{1+\Delta m} + \frac{\Delta m}{1+\Delta m} - H(z_a) \approx (1 - H(z_a))\,\Delta m.$$

Therefore, the resulting change in her position will be $\Delta z_a = -(1 - H(z_a))/h(z_a)\,\Delta m$. We can summarize the change in the two cases using the following two expressions: a constant negative term and a positive term that only operates for entrepreneurs E with ideas originally worse than z, as

$$\Delta z_a = \left[\frac{1}{h(z_a)}\mathbb{I}(z > z_a) - \frac{1 - H(z_a)}{h(z_a)}\right]\Delta m,$$

where $\mathbb{I}$ stands for an indicator function. We can think of the first term inside the bracket as whether or not z improves the pool of ideas for any given entrepreneur. The second term then is a constant crowding term that is identical for all firms z. Intuitively, adding any new firm makes it less likely for entrepreneurs to find their original ideas, and in this way the likelihood of adopting this new idea produces an opportunity (crowding) cost.[32]

Step (II): Accounting for the Expected Lifetime of Ideas The calculation above accounted for the quality of knowledge spillovers generated by different firms. Next, we have to account for

[32] Note that this force is also present in standard models of creative destruction (e.g., see Acemoglu et al., 2013).

the social values associated with these spillovers. The social values of ideas depend not only on their quality, but also on the time horizon over which they benefit the industry. From the social planner's perspective, each change Δz_a in the position of an entrant generates an external value $V'(z_e)\,\Delta z_a$, where the marginal value is calculated at the entry efficiency z_e which is related to the adopted efficiency z_a through $z_e \equiv \eta\tilde{z} + (1-\eta)\,z_a + \rho\gamma_e$.

Step (III): Summing Up the Effect Across All Entrants The two terms in Equations (30) and (31) sum up all of these external values across firms at different tiers of adopted efficiency z_a, accounting for the fact that the rate of adoption in each tier is given by $\lambda_e \times h(z_a)\,f_a(H(z_a))$.

The next corollary immediately follows from the proposition.

Corollary 1. *The market equilibrium is socially efficient if there are no spillovers, i.e. $\eta = 1$ and $\tilde{\theta}_t$ exogenous, or if no entry happens, i.e., $\lambda_e = 0$.*

When the distribution of productivity of new entrants is determined exogenously, market equilibrium innovation and production decisions become optimal. In the absence of spillovers, the efficiency results that Dhingra and Morrow (2012) have derived for the static benchmark model of firm heterogeneity and selection (Melitz, 2003) generalize to the dynamic case with firm innovation, which corresponds to the innovation model of Atkeson and Burstein (2010). Similarly, when the equilibrium path does not feature entry, knowledge spillovers do not play a role and the innovation and exit decisions of firms become socially optimal.

In addition to the characterization of the social value function, the proposition also provides the social free entry condition. Equation (33) shows that the optimal entry condition for the social planner takes the same form as the free entry conditions under the market equilibrium. While both the optimal distribution H_s and the value function V_s diverge to those under the market equilibrium, the relative value of entry (in terms of fixed operation costs) has to be the same between the two allocations.

3.2 Unpacking Innovation Externalities

Proposition 4 characterizes the socially optimal allocation. However, it does not directly show us how the market equilibrium and the socially optimal and innovation decisions deviate from each other for a firm in efficiency state z. These two investments satisfy

$$e^z J'(\gamma(z)) = V'(z)\,, \qquad e^z J'(\gamma_s(z)) = V_s'(z)\,, \tag{34}$$

respectively. Investigating the market equilibrium and social HJB equations, we find that the differences between the marginal private and social values of a firm with efficiency z derive from two distinct sources:

1. Knowledge spillovers: characterized by the endogenous distributions (H, H_s) and the exogenous adoption process F_a,
2. Expected lifetime (selection): characterized by the endogenous rates of productivity growth $(g_\theta, g_{\theta,s})$ and the exogenous volatility parameter σ.

To unpack these different sources, let us again use a perturbation argument. This time, consider raising the innovation investments of all firms within a Δz neighborhood of the efficiency level z by a small amount $\Delta\gamma$. I will examine the external impacts of this change that firms within Δz do not account for when they make their innovation decisions. The key is to note that, due to the dynamic effect of business stealing, other firms in the long-run equilibrium will respond to this change by lowering their own innovation investments. However, these long-run effects are fairly complex, since the initial change sets in motion a chain reaction of responses across different types of firms that propagates across the industry and creates a new equilibrium.

To isolate the main channels, I consider a partial-equilibrium setup whereby I consider only the *direct* effect of the change above on a group of firms at another efficiency level, say, x. Here, I assume that no other groups of firms respond except firms within a small neighborhood of x. Of course, one can then propagate these first-level responses to a second-level response, and so on. But the direct effect of this perturbation allows us to examine the main logic of the externalities.

If no other firms respond to the change $\Delta\gamma$ among firms z, the long-run rate of profit erosion rises to reflect the change. From the decomposition (26), we find this effect to be given by[33]

$$g \longrightarrow g + \nu(z) \times \Delta z \Delta\gamma,$$

where $\nu(z) = h(z)\, e^{z} J(z) / Z$ denotes the market share of firms with efficiency state z. We can identify the following types of externalities in the innovation response of all other firm.

Externality (I): Knowledge Improving When firms with efficiency z raise their investments, they raise their likelihood of moving up in the efficiency ladder. This raises the density of firms immediately above them, and accordingly changes the knowledge spillovers that entrants receive. Using the same argument as the previous section, we can see that the size of this change depends on the marginal social value of the resulting improvement in knowledge spillovers, and is proportional to $\lambda_e \times U_s'(z) \times h(z)\, \Delta z \Delta\gamma$.

Externality (II): Innovation Shifting However, the rise in the innovation investments of firms z raises the strength of competition across all the other firms. Other firms, say at efficiency state x, then respond to this competition by cutting down on their own innovation investments. This is the *innovation shifting* effect that we briefly discussed in the Introduction and earlier in this section. This effect is always negative since the innovation investments are decreasing in the rate of profit erosion, $\partial\gamma/\partial g < 0$, and the spillover function is monotonically increasing $U_s'(x) > 0$. Figure 3 provides a graphical illustration of this dynamic displacement effect. It shows that, starting from the dashed, market-equilibrium schedule of innovation investments as a function of efficiency states, we push the investments of firms in the band Δz around z upward by $\Delta\gamma$ to find the solid schedule. For firms at efficiency state x, the negative impact is proportional to $U_s'(x)$. The total innovation shifting effect of the rise in innovation of firms around z corresponds to the integral of

[33] I emphasize that this expression only captures the direct effect of the innovation, under the assumption that other incumbents and entrants do not change their strategies immediately. It is certainly possible that the general equilibrium response of other firms results in the fall of the rate of profit erosion in response to the perturbation considered here.

this effect across all firms x. Crucially, due to the higher market shares of productive firms, this negative effect rises with efficiency of the original band of firms around z.

Externality (III): Stifling Effect on Startups Competition also affects the incentives for new firms to enter the market. The faster rate of growth changes the expected value of entry and therefore potential entrants respond by reducing the rate of entry. This results in a reduction in spillovers proportional to $\mathbb{E}_{F_s}\left[U_s\left(z\right)\right] - \Psi_s$, where F_s denotes the distribution of entrant efficiency.

Externality (IV): Cleansing Effect on Unproductive Firms Finally, competition raises the rate of exit among the least productive firms. This stronger selection lowers the negative crowding costs on the industry, proportional to Ψ_s. This is a positive effect. Figure 3 shows a graphical illustration of this effect for firms near the exit threshold.

The net external effect of the rise in the innovation investment of firms z depends on all these components. However, the industry response is not limited to these; they are only the immediate impacts of the perturbation that clarify the channels. These effects in turn trigger further innovation and exit and entry responses among other firms. In the long-run the stationary distribution H also responds to the perturbation. The combination of the responses of other firms and the stationary distribution may in fact give rise to a situation where the aggregate rate of productivity growth falls as the result of this perturbation.

Where does the expected lifetime of the ideas come into the picture? It manifests itself in the relationship between social knowledge spillover $U_s\left(z\right)$ and efficiency states z. Under standard theories of creative destruction, all ideas have the same horizon of usefulness in the market. From the social planner's perspective, this horizon is infinity. Therefore, the only margin for variations in the externalities is the impact of ideas in generating a flow of profits. We find that both spillover externalities and the displacement effects discussed above scale proportionally to the market share of firms. As a result, the gap between the two remains constant across firms. In contrast, in this model the displacement externalities grow as the market share of firms (which grows at e^z) whereas the knowledge spillovers grow at $U_s\left(z\right)$ (which grows faster than e^z for firms near the exit threshold). As a result, the gap between the two may vary across firms as a function of efficiency states.

3.3 Computing the Socially Optimal Allocation

Despite the complex and interconnected nature of the displacement effects, we can still numerically compute the optimal allocation and compare it with the market equilibrium one. This section discusses a scheme that allows us to solve for the optimal allocation.

To solve for the optimal allocation, I use a computational scheme that parallels the one used for solving the market equilibrium. However, the key difference here is that the industry and incumbent problems are fully entangled through the term involving the spillover function in Equation (29). Therefore, in order to solve for the value function, investment, and exit policy functions of firms, we need to know the stationary distribution of firm efficiencies.

I use an iterative scheme to solve the two problems jointly. Starting with an initial guess for the rate of profit erosion, I solve for the value function of the firm and the stationary distribution of firm efficiencies, under the market equilibrium. Next, I use these results to form a guess about the spillover function and solve for the social value function from Equation (29). Then, I continue to iteratively solve the social HJB and KFE until convergence. I use the difference between the resulting cost of entry and the actual cost of entry to update my guess about the optimal rate of profit erosion and repeat the same procedure. The final solution satisfies all the different conditions in Proposition 4.

3.4 Tax/Subsidy Implementation of Optimal Policies

In this section, I discuss potential implementations of the optimal allocation under the assumption of information symmetry between the government and private firms. Since the government can observe the productivity ranking of firms (efficiency z), we can reach the first-best through two different potential schemes. First, the government can simply offer an efficiency-dependent reward schedule $u_s(\cdot)$ and a lump-sum tax of ψ_s to firms. This scheme directly implements the socially optimal allocation and does not require any further incentives for entry. We can think of this scheme as an *ex post* reward that corrects for the gap between social and private returns to innovation investments. Needless to say, this scheme is fairly costly, and involves large transfers between the government and private businesses.

Alternatively, the government can use a combination of targeted taxes/subsidies to implement the optimal allocations characterized in Section 3.1. In particular, the following set of instruments together achieve the optimal allocation: a flat tax on the operation of firms, a subsidy for firm startup costs, and two additional efficiency-dependent instruments, a subsidy on innovation investments and a tax on the labor inputs of incumbents. The lump-sum tax internalizes the negative crowding externality of firms and the entry subsidy internalizes the entrants' positive spillovers. The efficiency-dependent input taxes and subsidies together internalize innovation externalities while maintaining the static efficiency of allocations across firms.

To see how this works, consider a firm with efficiency z that invests the amount $\gamma_s(z)$ under the optimal allocation. The government offers tax $\tau_l(z)$ on production labor inputs and subsidy $\tau_i(z)$ to the innovation labor inputs of the firm. The purpose of the production input tax is simply to ensure the static efficiency of firm outputs: therefore the government chooses

$$\tau_l(z) = -\tau_i(z)\ \varphi(\gamma_s(z)), \tag{35}$$

to ensure that the total labor costs of the firm under the implementation scheme match those under the optimal allocation. In addition, because of this tax the government does not need to transfer a net amount to the firm.

Now, if we let τ_o denote the tax on the operation of firms, the value function of the firm under the implementation satisfies the following HJB equation

$$r\,\widehat{V}_s(z) = \max_{\gamma} e^z J_\tau(\gamma; \tau_i(z), z) - (1+\tau_o) + \rho(\gamma - g_s)\,\widehat{V}_s'(z) + \frac{1}{2}\sigma^2\,\widehat{V}_s''(z)$$

where I have defined the function

$$J_\tau(\gamma; \tau_i, z) \equiv (1 + (1+\tau_i)\,\varphi(\gamma) - \tau_i \varphi(\gamma_s(z)))^{-\rho}.$$

It is easy to show that the derivative of this function with respect to γ, when evaluated at $\gamma_s(z)$, is given by $(1+\tau_i)\,J'(\gamma_s(z))$. Together with condition (34), this implies that if we set

$$\tau_i(z) = \frac{\widehat{V}_s'(z)}{V_s'(z)} - 1,$$

where the social value function V_s satisfies the social HJB (29), the choice of innovation of firm z under the implementation coincides with its choice under the optimal policy.

The following lemma then characterizes the set of instruments that implements the optimal allocation.

Lemma 1. *(Tax/Subsidy Implementation) The social planner can achieve the optimal allocation by imposing a flat tax τ_o on the operation of firms, a tax/subsidy $\tau_{e,s}$ on the costs of entry, and a size-dependent tax/subsidy $\tau_{i,s}(\cdot)$ on incumbent innovation investments. Let $\widehat{V}_s(\cdot)$ denote the value function of the firm under the market equilibrium featuring the additional taxes and subsidies. This value function and the operation tax τ_o satisfy the following differential equation*

$$r\,\widehat{V}_s(z) = e^z J(\gamma_s(z)) - (1+\tau_o) + \rho\,(\gamma_s(z) - g_s)\,\widehat{V}_s'(z) + \frac{\rho^2}{2}\sigma^2\,\widehat{V}_s''(z), \tag{36}$$

subject to the boundary conditions $\widehat{V}_s'(z_{o,s}) = \widehat{V}_s(z_{o,s}) = 0$, where $(\gamma_s(\cdot), z_{o,s})$ stand for the firm policies under the optimal allocation from Proposition 4. Given the value function $\widehat{V}_s(\cdot)$ and the social value function V_s, satisfying social HJB (29), the incumbent and entrant innovation taxes/subsidies that implement the optimal allocation are given by

$$\tau_{i,s}(z) = \frac{\widehat{V}_s'(z)}{V_s'(z)} - 1, \tag{37}$$

$$\tau_{e,s} = \frac{\mathbb{E}_{F_s}\left[\widehat{V}_s'(z)\right]}{\mathbb{E}_{F_s}[V_s(z)]} - 1. \tag{38}$$

Note that Equation (36) has a different form from the social planner's HJB equation (29). As a result, the social value function of the firms V_s and their private value under the subsidy/tax implementation $\widehat{V}_s$ do not necessarily coincide. Therefore, the subsidies in Equation (37) and (38) will in general be nonzero.

4 Quantitative Results

In this section, I use data on the life cycle dynamics of firms to calibrate the parameters of the model and to quantitatively investigate its positive and normative predictions. The model is parsimonious and characterizes each sector of the economy by three key sets of structural characteristics relating to 1) process of innovation (volatility, and costs of innovation and entry), 2) demand (elasticity of substitution and the rate of growth of demand), and 2) the process of entry and adoption (directedness and the degree of decreasing returns).

I face a challenge in calibrating the parameters of the process of entry, since reliable measures of the productivity of entrants are rarely available. Employments, sales, or even estimates of multifactor productivity do not provide precise measures for the performance or efficiency of new firms. In the early stages of their life cycles, firms invest more heavily in tangible and intangible capital relative to mature firms and therefore the comparison between the two groups may not be justifiable. As I will explain below, I rely on the relationship between the rate of exit and the age of the firm to identify the quality of entering cohorts relative to other firms active in the industry.

4.1 Data

I use the Business Dynamics Statistics (BDS) tables, a dataset that has been specifically developed to account and track sources of job creation and destruction through entry and exit, disaggregated by firm size, age, and industry.[34] The data starts in 1977, but the truncation of the firm age variable is heavy for the first few years. I focus on the period 1987-2007 to allow for a broader coverage of the age variable and avoid the arguably different firm dynamics during the Great Recession.

As I will explain below, the empirical exercise further requires the rates of growth in the aggregate employment and labor productivity of firms in each sector. BDS data already provides information on the total employment of firms. For measures of sectoral productivity growth, I rely on the Bureau of Labor Statistics (BLS) National Accounts to obtain estimates of the average rate of labor productivity growth in each sector for the period under study.[35]

4.2 Model Specification

The model in Section 2 characterizes a single-sector general equilibrium economy, but I perform the calibration at the level of two broad sectors of the US economy: manufacturing (SIC code 20) and retail trade (SIC code 52). I sketch a two-sector generalization of the model in Section 7.3 of the Appendix, and show that the model is compatible with an equilibrium path along which each sector grows in terms of average productivity and mass of firms with a sector-specific rate. This provides a theoretical ground for separately calibrating the model to these two sectors of the US economy.[36]

I assume the following form for the costs of innovation

$$\varphi(\gamma) = \chi \frac{\gamma^{1+\kappa}}{1+\kappa}, \tag{39}$$

[34]The underlying data for LBD and BDS is at the establishment level, but the Census Bureau matches establishments based on the commonality of ownership or control to create firm-level data (see Haltiwanger et al., 2009). The dataset can be accessed online at http://www.census.gov/ces/dataproducts/bds/.

[35]BLS estimates sectoral labor productivity based on measures of value added and hours worked provided in the NIPA accounts (for the details of the methods used, see Bureau of Labor Statistics, 2008). The dataset can be accessed online at https://www.bls.gov/lpc/.

[36]There is evidence in the data that the sectors in fact have been undergoing some restructuring and therefore may not be precisely evolving along their long-run stationary equilibrium paths. However, the focus of this work is in documenting the essential inter-sector differences in the mechanics of growth. Hence, I abstract away from characterizing the transitional dynamics, and assume that the two sectors observed in the data evolve along their respective stationary balanced growth paths and match the data moments with those implied by the model.

defined in the interval $\gamma \in [0, \overline{\gamma}]$, where

$$\overline{\gamma} \equiv \left(\frac{\kappa}{\chi} \frac{1+\kappa}{\rho(1+\kappa)+1} \right)^{\frac{1}{1+\kappa}}. \tag{40}$$

For the process of diffusion, I consider the adoption distribution (11) that allows for a potential bias of the distribution toward more productive firms. In addition, I assume the relation (13) between entry and adopted productivities, and let $\tilde{\theta}_t$ denote the average productivity of the industry at time t. Then, the tuple (μ, η, γ_e) characterizes the parameters of the adoption process. See Section 7.3 of the Appendix for more details.

4.3 Calibration Strategy

To calibrate the main parameters of the model, I find the values that match the predictions of the model with the data on the sectoral rates of growth and the life-cycle dynamics of firm employment.

For the parameters of demand, I assume a discount rate $r = 0.05$ and an elasticity of substitution of $\rho = 2$ for both sectors, both values common in the literature.[37] As I show in Section 7.3 of the Appendix, the model allows me to separately choose the growth rates g_N, corresponding to both rates of growth in the number of firms and in total employment in the two sectors. I choose to match this value to the rate of growth of total sector employment based on the BDS data.

The remaining parameters of the model are (1) the primitives of the innovation process, i.e., relative costs of entry ψ, size and curvature of the innovation cost function (χ, κ), and volatility σ, and (2) the primitives of the diffusion process, i.e., the directedness μ, the decreasing returns parameter η and the shifter γ_e. I choose to search in the space of 5 parameters $(\chi, \sigma, \mu, \eta, \gamma_e)$, and to directly use two moments from the data and prior work to pin down the two parameters (ψ, κ).[38]

To find the two parameters (ψ, κ), I employ the following scheme. Given other model parameters, I find the curvature of the innovation cost function κ to match the values reported in the empirical literature for the user cost elasticity of firm-level R&D spending.[39] The literature typically relies on variations in the corporate tax code over time (Hall, 1993; Hall and Van Reenen, 2000), across countries (Bloom et al., 2002), and across US states (Wilson, 2009) as sources of changes in user costs of R&D to estimate the response in the R&D expenditure of firms. I follow Akcigit and Kerr (2012) and Acemoglu et al. (2013) who conclude that an elasticity of around -1 well summarizes the findings of this literature. Therefore, I set κ such that the corresponding value in the model is -1. Finally, the zero entry condition in Equation (23) suggests that given other model parameters, the rate of productivity growth g_θ pins down the relative costs of entry ψ in each

[37] The choice of $\rho = 2$ implies an elasticity of substitution of 3, close to the average of the estimates that Broda and Weinstein (2006) find for the elasticity of substitution based on US import demand data. I choose this number, following previous calibration exercises that rely on CES industry-level aggregators to study within-industry firm heterogeneity (e.g., Hsieh and Klenow, 2009; Perla et al., 2015).

[38] The fixed cost of operation ψ_f does not play a role in determining the dynamics of the model in the long run.

[39] In Section 7.3 of the Appendix, I derive the relation between the parameter κ and the observed user cost elasticity of innovation investments based on the model. Using this relation, the latter directly pins down the value of the former, given other parameters of demand and innovation environment, and the rate of sectoral productivity growth.

sector.

I search for the set of parameters $(\chi, \sigma, \mu, \eta, \gamma_e)$ that minimize the distance between the moments observed in the data and the ones predicted from the model (for previous examples of a similar approach, see Lentz and Mortensen, 2008; Acemoglu et al., 2013). I target the following three sets of moments:

1. Rate of entry,
2. Rate of employment growth and reallocation for old (15-20 year) firms,
3. Rates of exit for mature (10-15 year) and old (15-20 year) firms.

The first three moments (1 and 2) help identify the costs of innovation χ , the extent of volatility σ, and the directedness of the adoption process μ. The third group of moments together provide us with information about the relationship between age and the rate of exit in the industry, which informs the identification of the decreasing returns parameter η. If entrants begin with low-quality ideas relative to incumbents, they exit heavily in their early stages of life since they are very likely to cross the exit threshold early on due to negative shocks. If entrants begin with high-quality ideas relative to incumbents, their rates of exit has a more gradual relationship with age, since enough negative shocks have to accumulate in order for them to cross the exit threshold. This allows us to infer the initial characteristics of firms based on their likelihood of exit as a function of age.

Let $\mathbf{p}$ denote the vector of model parameters and $\widehat{\mathbf{m}}$ a vector of targeted empirical moments observed in the data. Let $\mathbf{m}(\mathbf{p})$ denote the vector of the expected values for the same moments based on the model, when calibrated with the set of parameters $\mathbf{p}$. To find the set of parameters that best explain the data, I use the following criterion

$$\mathbf{p}^* \equiv \min_{\mathbf{p}:\, \mathbf{p}_{fix} = \mathbf{p}^*_{fix}} \|\mathbf{m}(\mathbf{p}) - \widehat{\mathbf{m}}\|_{\widehat{\Sigma}}, \tag{41}$$

where $\mathbf{p}_{fix}$ is the subset of parameters fixed to $\mathbf{p}^*_{fix}$ (see above), and $\widehat{\mathbf{\Sigma}}$ is a positive definite matrix computed based on the variance of the moments in the data over time. We cannot derive analytical expressions for moment functions $\mathbf{m}(\mathbf{p})$ under the model. Hence, I use simulation to compute the predictions of the model for each moment.[40] Finally, I employ a simple brute force grid search to minimize the objective function of the problem (41). I search in a subset of the parameter space that yields a feasible solution and find the parameters with the lowest value for the objective function (41).

4.4 Calibration Results

Table 1 presents the calibrated moments for the two sectors. The top three moments are each directly used in the calibration of one specific parameter. Therefore, the model and the data exactly

[40] The logic of the approach is the same as that of the Method of Simulated Moments (MSM) for estimation (McFadden, 1989; Pakes and Pollard, 1989; Duffie and Singleton, 1993; Gourieroux and Monfort, 1993). The key challenge in computing the moments implied by the model stems from the continuous-time nature of the underlying stochastic process. A wide literature in empirical finance has studied the problem of estimating the parameters of continuous-time stochastic differential equations such as those used in this paper using simulation-based/analytical and ML/GMM-based methods (Pedersen, 1995; Aït-Sahalia, 2002; Durham and Gallant, 2002; Brandt and Santa-Clara, 2002). Here, I use a simple and fast simulation-based approach that relies on time discretization and importance sampling (Pedersen, 1995).

match with regard to these moments. The case of the rate of employment growth in manufacturing poses an exception, since the data suggests a negative rate, but the model requires a positive rate of growth of employment to ensure the existence of a stationary long-run equilibrium.[41] Therefore, I instead assign a very small positive value for g_N in manufacturing (equal to1×10^{-4}) to deliver the existence of a long-run equilibrium. The remaining five moments are used for the grid search according to the objective function in Equation (41).

The calibrated parameters show some differences between the two sectors. As one would expect, the relative costs of innovation are higher in retail compared to manufacturing. The volatility parameter is higher in retail due, among other things, to the higher rate of reallocation of employment for old firms. This is in line with the facts documented by previous empirical work (Foster et al., 2006). The parameters of the process of entry and adoption (μ, η) also present clear contrasts between the two sectors. The decreasing returns parameter appears higher in retail trade. The calibrated parameters suggest that firms enter the market with an efficiency (log productivity) that is only 40 percent that of the firms they are trying to imitate (70 percent in manufacturing), and have to build up the remainder of their competency by investing in innovation through their life cycle.[42] Whereas the adoption seems fairly directed toward high-productivity firms in retail, uniform adoption provides the best fit in manufacturing. In other words, entrants are more likely to adopt the ideas of the most productive incumbents in retail. Together these calibrated parameters suggest that the adoption process may be more directed but less perfect in retail compared to manufacturing.[43]

Targeted Moments The model provides a good fit for the targeted moments, and in particular those that capture the relationship between age and the likelihood of exit. In the cases of the rates of growth and reallocation of old firms, the fit is not as close. The reason is that the data in fact suggests a strong trend in the relationship between age and volatility, which the current simple model is abstracting from.[44] This means that the calibration has to find a compromise in fitting the patterns for old versus young firms. Nevertheless, the calibration preserves the ordering of the moments between retail and manufacturing for all the moments.

[41] The leading explanation for a negative trend in the data is that the sector is moving along an adjustment path toward a new stationary equilibrium with a smaller manufacturing sector. This could be due to the changes in the technological environment (e.g., skill-biased technical change) or the effects of globalization (rise of the Japanese manufacturers in the 70s and 80s and Chinese manufacturers in the 2000s).

[42] The model infers this mainly based on the fact that the rate of exit of mature and old firms relative to overall the rate of turnover, as proxied by the rate of entry, is higher in manufacturing than in retail. One way to interpret this is that the retail sector is better characterized by a so-called *revolving-door* feature, in which most entrants soon exit and rarely grow to penetrate into the small band of industry leaders. Audretsch (1995) introduces this concept and contrasts it with what he calls a *forest* industry where leading incumbents are frequently taken over by new entrants that slowly grow from beneath. Seen from the prism of the current model, the differences between the two evolutionary modes reflect the differences in the relative position of entrant ideas, relative to the incumbents they are trying to imitate.

[43] This finding is compatible with the view that an important component of the competency of a firm in retail trade corresponds to the types of intangible capital such as building supplier networks, branding, organizational capacity that the firm has to acquire throughout its life-cycle, and cannot simply imitate from others. For instance, a more detailed structural account of the differences between retail and manufacturing should take into account the process of market penetration and spatial expansion, which inevitably requires an extended catch-up process for entrants (for the example of the growth of Walmart, see Holmes, 2011).

[44] Interestingly, the data suggests that the gap in the volatility of young and old firms is higher in manufacturing relative to retail, even though on average the volatility is lower in the former.

Growth Decomposition and Untargeted Moments Table 2 presents a number of different moments predicted by the model under the calibrated parameters for the two sectors (in the columns under the heading "Before").

The first three rows show the decomposition of Equation (26), separately showing the contributions of turnover (replacement of exiting firms with entrants) and net entry. First, note that, relative to the overall sectoral rate of productivity growth, the contribution of turnover is considerably larger in retail compared to manufacturing. This is the case despite the fact that, as the fourth line in the table shows, the gap in the productivity of entering and exiting firms is smaller in retail compared to manufacturing. The faster rate of turnover in retail compensates for this smaller gap and results in a stronger contribution of selection to growth in this sector. However, the third row indicates another difference between the two sectors in the contribution of entry: faster growth in the number of retail firms means that the sector is constantly flooded by small and relatively unproductive entrants. The negative contribution of this component more than cancels out the contribution of selection.[45]

The sixth row shows that the Pareto tail indices in both sectors are very close to one, and therefore suggest the approximate emergence of the Zipf law (Gabaix, 2016). This contrasts with the workhorse models of innovation and growth (Klette and Kortum, 2004; Lentz and Mortensen, 2008) that predict counterfactually thin tails for the distribution of firm size (see Appendix 6 for further discussion). Relatedly, the stationary distribution of efficiencies in manufacturing exhibits a larger dispersion of productivity compared to retail. The interdecile range (IDR) of log productivity is 2.6 in manufacturing and 2.0 in retail. This is compatible with the findings of Foster et al. (2006), who highlight the larger dispersion of productivity in manufacturing despite the higher volatility of businesses in retail.

Table 2 also compares the relative dispersion of innovation investments between the two sectors. As the measure of this dispersion, I compute the interdecile range (IDR) of the investments, normalized by the maximum level of investments in each sector. The interdecile range of the investments in retail is predicted to be slightly larger than that in manufacturing. The model makes this prediction based on the higher volatility of firm outcomes in retail compared to manufacturing. In a volatile environment, highly productive firms have a stronger advantage in process innovation because they expect to reap the benefits of their investments over a longer horizon. High productivity shields these firms from the threat of imminent exit, which undermines the innovation incentives of unproductive firms in a volatile industry.

Counterfactual Experiment: Reduction in the Costs of Innovation Next, I use the model to explore the consequences of a large reduction in the costs of innovation in the two sectors. I study what happens when due to forces external to the sector, for instance, the emergence of novel General Purpose Technologies (GPTs), new opportunities for innovation become available

[45] The closest empirical counterparts to these results are the productivity decompositions reported by Foster et al. (2001) and Foster et al. (2006) in the two sectors. They find a far more pronounced difference in the contribution of entry and exit to growth in between the two sectors, with entry and exit explaining almost all of the growth in retail in the 90s while accounting for about 25% of growth in manufacturing within the same period. They also find a larger gap between the productivity of entering and exiting establishments in retail compared to manufacturing. However, note that Foster et al. (2006) focus mainly on entry and exit of establishment rather than firms. It is likely that their results are driven by the entry of establishments that belong to high-productivity incumbent firms, fast growing retail chains such as Walmart and Target, rather than new firms.

to all firms. I study this question by examining the changes in the long-run equilibrium when the parameter of innovation costs χ falls to 75 percent of its original level in both sectors. I assume all other parameters, including the costs of entry, remain the same before and after this change.

Table 2 shows a number of different outcomes before and after the reduction in costs. The long-run rate of productivity growth in both sectors rises by about 0.3 percent in response to this change. The lion's share of this response is provided by the rising investments of incumbent firms as we find a very small response in the contribution of turnover. Interestingly, the impact on the contribution of turnover to growth is in reverse direction between the two sectors. The rate of entry mildly rises in response to lower costs in both sectors. This response is mainly driven by the responsiveness of the exit decisions of unproductive firms to the higher innovation investments of productive firms. Finally, the overall impact of these changes raises the concentration in both industries, as reflected in the reduction in the tail index of the long-run distribution. The sectors become more concentrated, since more productive firms disproportionately benefit from the lower innovation costs, while the costs of entry have remained intact. This is compatible with the evidence on the rise in the concentration of the retail sector (Foster et al., 2006).

4.5 Optimal Allocations

Let us now examine the optimal allocation of innovation investments and production in the two sectors. These allocations solve the social planner's HJB equation (29), the corresponding KFE, and the social free entry condition (33).

Figure 4 compares the market equilibrium and socially optimal innovation investments in the two sectors. Figure 5 performs the same comparison, but for the long-run distribution of productivity in the two sectors. Finally, Figure 6 shows the difference between the market equilibrium and the socially optimal allocation for the two variables.

First, observe that in both sectors, large and productive firms overinvest in innovation. Relative to the innovation intensity of each sector, the extent of misallocation in the innovation investments appears somewhat larger in retail compared to manufacturing. Examining the differences in innovation investments Figure 6, we find that underinvestment of firms in the medium range of productivity is particularly severe in retail trade. For a range of medium efficiency firms in retail, the socially optimal innovation investments reaches the maximum value given by Equation (40) and is constant. This suggests a strong negative displacement effect stemming from the innovation investments of high-efficiency firms in this sector.

Under the socially optimal allocation, the long-run distributions of efficiency in both sectors shift to the left. This means that the concentration of both sectors under the market equilibrium is suboptimally high. The socially optimal allocation raises the rate of productivity growth and therefore the rate of dynamism of the industry and the obsolescence of ideas. As a result, it is harder for firms to grow very large under the optimal allocation.

Finally, note that Figure 6 presents the results as a function of efficiency relative to the cutoff under the corresponding allocation. The socially optimal exit cutoffs are larger than those of the market equilibrium because the social planner internalizes the negative crowding externalities of firms on entrants.

Implementation of the Optimal Allocation I use the scheme introduced in Lemma 1 to compute the set of subsidies and taxes that the social planner can employ to implement the optimal policy under the market equilibrium. These Pigouvian taxes serve the purpose of showing us the wedges that exist under the market equilibrium between the social and private returns to innovation investments. Table 3 compares this policy between the two sectors. The average subsidy rate that the scheme offers to incumbents and entrants is higher in retail trade compared to manufacturing. Figure 7 shows that the profiles of the schedule of subsidies as a function of firm efficiency also appear fairly different between the two sectors.

First, note that the relationship between efficiency and the subsidy rate is nonmonotonic. The subsidy is increasing with efficiency among the lower tiers of productivity, but then declines among high-efficiency firms. However, the initial rise is much less pronounced in retail compared to manufacturing. Figure 8 shows that this increasing relationship can be well approximated with a linearly growing subsidy rate for a vast majority of firms (for firms up to the 74 percentile of productivity and size in retail, and 84 in manufacturing). In Table 3 (row 7), we can compare the slopes of the approximate schedule of subsidy rates between the two sectors. Among lower-productivity firms, the relationship between the subsidy rate and productivity is far weaker in retail compared to manufacturing. In contrast, among high-efficiency firms, the relation between productivity and the optimal subsidy rate is stronger in retail: the subsidy declines faster among high-efficiency firms in this sector. This point relates to the observation we made earlier: the negative displacement effect of high-productivity firms is relatively larger in retail compared to manufacturing.

Comparison with Uniform Incumbent Subsidies How much do we gain by implementing the nonlinear subsidy scheme above, relative to a uniform one-size-fits-all subsidy rate to the innovation investments of all incumbents? To investigate this question, I take the following strategy. I compare the amount of spending by the government on subsidies to entrants, under the subsidy schedule that implements the optimal allocation and under a uniform subsidy scheme that minimizes spending. As with the optimal subsidy, I assume that the government in the case of uniform subsidies also imposes a tax on the wages of production workers, given by Equation (35), to cancel out the effect of subsidies on the production decisions of firms. As a result, the net spending of the government on subsidies is given by the amount paid to entrants.

For each constant rate of uniform subsidies, I find the corresponding subsidy rate to entrants that achieves the same rate of productivity growth as that of the optimal allocation. To control for the effect of lump-sum operation taxes, I perform the comparison with and without the additional lump-sum tax for the case of the uniform subsidies.[46]

The last two rows of Table 3 show considerable gaps between the spending on subsidies between the scheme that optimally accounts for heterogeneity and one that ignores it. I conclude that the model implies misallocations in the innovation investments of firms that are first-order to the design of policies that aim to internalize the externalities of innovation.

[46] This scheme effectively solves the constrained optimum subsidy problem, where the imposed constraint is the constancy of the subsidy rates for incumbent firms.

4.6 Effect of Static Competition

As we saw, the model implies sizable misallocations in innovation investments across heterogeneous firms. How does the strength of product market rivalry, as proxied by the degree of substitutability between firm products, affect these misallocations? To investigate this point, I perform the following exercise. I consider the effect of changing the elasticity of substitution parameter from $\rho = 2$ to $\rho_{new} = 3$. In each sector, I adjust the parameters of the costs of innovation (χ, κ) and entry ψ such that the cost elasticity of innovation, the rate of growth of large firms, given by $\rho\left(\gamma^{**} - g_\theta + \frac{1}{2}(\rho - 1)\sigma^2\right)$, and the rate of productivity growth remain the same. However, I keep the volatility parameter σ and the parameters of the process of knowledge diffusion (η, γ, γ_e) intact, so that the analysis only focuses on the effect of static product market rivalry.

Figure 10 shows the difference between the market and socially optimal investments in innovation for the two sets of parameters with the low and high substitutability. Figure 9 compares the optimal subsidy schemes between the two product market competition settings. Based on the discussions above, we should expect that stronger competition amplifies the negative externalities of large firms on smaller ones. As a result, with stronger competition, the peak of the gap in the market and optimal investments, and the subsidy schemes in both sectors shifts to the left.

5 Discussion and Future Work

Knowledge spillovers are the cornerstone of our modern theories of growth. In this paper, I study them in a standard model of firm heterogeneity and selection (Hopenhayn, 1992; Melitz, 2003) where firms actively invest in innovation (Ericson and Pakes, 1995; Atkeson and Burstein, 2010). I find that, under a fairly broad set of assumptions, markets misallocate innovation investment across firms. In a calibration based on data from two broad sectors of the US economy, I find that these misallocations are quantitatively large and removing them could result in a substantial rise in the aggregate rate of productivity growth.

The model offered here only focuses on the dynamic effects of competition that work through the expected lifetimes of firms. Specifically, the assumption of CES demand implies constant markups that do not respond to the innovation investments or changes in firm productivity. More importantly, under this assumption, the extent to which firms cannot fully appropriate the social value of innovation exactly equates their business stealing effect on other firms. However, extensions of this model can accommodate more general demand systems such as those with Variable Elasticities of Substitution (VES) in which both markups and the gap between appropriability and business stealing effects vary across firms (Zhelobodko et al., 2012; Dhingra and Morrow, 2012) .

In this paper, I mainly characterize the behavior of industry equilibria along the stationary constant growth paths. Therefore, I do not examine the regularities observed along the industry life cycle, which has been the focus of a separate line of theoretical and empirical work. Klepper and Graddy (1990) and Klepper and Simons (2000) document the fact that as a new industry emerges, the number of firms rises, only to fall again after a so-called shake-out phase. Among others, Jovanovic and MacDonald (1994c) and Klepper (1996) offer theoretical frameworks that explain these patterns. I note that some of these patterns may also be rationalized through the model offered in this paper along the transitional dynamics of the path of the industry. I leave the investigation of this point for future research.

To my knowledge, this paper is the first to study potential inter-industry variations in the mechanics of productivity growth through the lens of the theories of firm dynamics. The empirical evidence suggests that the more prominent role of entry and exit in retail, compared to manufacturing, may be representative of a broader pattern across service-based industries (Foster et al., 2001). The growing role of service-based industries in employment both in the US and across the world requires us to understand the potential policy implications of these differences. By enriching the account of volatility and spillovers, this paper takes the first steps toward understanding the structural features that determine both the design of innovation policy and the response of growth to other interventions.

References

Acemoglu, Daron and Dan Cao, "Innovation by Entrants and Incumbents," *Journal of Economic Theory*, 2015, *157*, 255–294. 15

__ , Philippe Aghion, and Fabrizio Zilibotti, "Distance to Frontier, Selection, and Economic Growth," *Journal of the European Economic Association*, jul 2006, *4* (1), 37–74. 6

__ , Ufuk Akcigit, and Murat A Celik, "Young , Restless and Creative: Openness to Disruption and Creative Innovations," 2014. 52

__ , __ , Nicholas Bloom, and William Kerr, "Innovation, Reallocation and Growth," 2013. 2, 7, 23, 30, 31, 53

Aghion, Philippe and Peter Howitt, "A model of growth through creative destruction," 1992. 2

Aït-Sahalia, Yacine, "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach," *Econometrica*, 2002, *70* (1), 223–262. 31

Aizcorbe, Am, "Toward better measurement of innovation and intangibles," *Survey of Current Business*, 2009, (January), 10–24. 12

Akcigit, Ufuk and William R Kerr, "Growth through Heterogeneous Innovations," 2012. 30, 52, 53

__ , Douglas Hanley, and Stefanie Stantcheva, "Optimal Taxation and R&D Policies," 2016. 2, 5

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas, "Models of Idea Flows," 2008. 6

Aoyama, Hideaki, Hiroshi Yoshikawa, Hiroshi Iyetomi, and Yoshi Fujiwara, "Productivity dispersion: Facts, theory, and implications," *Journal of Economic Interaction and Coordination*, 2010, *5* (1), 27–54. 52

Arkolakis, Costas, "A Unified theory of firm selection and growth," *Quarterly Journal of Economics*, 2016. 12, 52

Atkeson, Andrew and Ariel T Burstein, "Innovation, Firm Dynamics , and International Trade," *Journal of Political Economy*, 2010, *118* (3), 433–484. 3, 8, 10, 12, 24, 36

__ and __ , "Aggregate implications of innovation policy," 2015. 2, 6, 12

Audretsch, David B, *Innovation and Industry Evolution*, MIT Press, 1995. 32

Aw, Bee Yan, Mark J Roberts, and Daniel Yi Xu, "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 2011, *101* (June), 1312–1344. 6

Axtell, Robert L., "Zipf distribution of U.S. firm sizes," *Science*, 2001, *293* (5536), 1818–1820. 4, 52

Bartelsman, Eric and John Haltiwanger, *Measuring and Analyzing Cross-Country Differences in Firm Dynamics* number January 2009. 2

_ , _ , and Stefano Scarpetta, "Cross-Country Differences in Productivity: The Role of Allocation and Selection," *American Economic Review*, feb 2013, *103* (1), 305–334. 52

Basker, Emek, "The causes and consequences of Wal-Mart's growth," *Journal of Economic Perspectivess*, 2007, *21* (3), 177–198. 3

_ , "The evolution of technology in the retail sector," in "Handbook of the Economics of Retailing and Distribution," Edward Elgar Publishing, 2016, pp. 38–53. 5

Benhabib, Jess, Christopher Tonetti, and Jesse Perla, "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier," 2014. 6

Bloom, Nicholas and John Van Reenen, "Why do management practices differ across firms and countries?," *Journal of Economic Perspectives*, 2010, *24* (1), 203–224. 10

_ , Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4), 1347–1393. 2

Bloom, Nick, Rachel Griffith, and John Van Reenen, "Do R&D tax credits work? Evidence from a panel of countries 1979-1997," *Journal of Public Economics*, jul 2002, *85* (1), 1–31. 30

Brandt, Michael W. and Pedro Santa-Clara, "Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets," *Journal of Financial Economics*, 2002, *63* (2), 161–210. 31

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121* (2), 541–585. 30

Bureau of Labor Statistics, "Technical Information About the BLS Major Sector Productivity and Costs Measures," Technical Report 2008. 29

Carvalho, Vasco M. and Basile Grassi, "Large Firm Dynamics and the Business Cycle," 2015, pp. 1–64. 54

Chaney, Thomas, "Distorted gravity: The intensive and extensive margins of international trade," *American Economic Review*, 2008, *98* (4), 1707–1721. 52

Cohen, Wesley M and Steven Klepper, "A Reprise of Size and R&D," *Economic Journal*, 1996, *106* (437), 925–951. 52

Comin, Diego and Martí Mestieri, *Technology Diffusion: Measurement, Causes, and Consequences*, Vol. 2, Elsevier B.V., 2014. 6

_ , Danial Lashkari, and Marti Mestieri, "Structural change with long-run income and price effects," 2015. 67

Corrado, Carol A. and Charles R. Hulten, "How do you measure a "technological revolution"?," *American Economic Review*, 2010, *100* (2), 99–104. 12

— , — , **and Daniel Sichel**, "Intangible Capital and U.S. Economic Growth," *Review of Income and Wealth*, 2009, *55* (3), 661–685. 12

Davis, Steven J., John Haltiwanger, and Scott Schuh, "Small business and job creation: Dissecting the myth and reassessing the facts," *Small Business Economics*, 1996, *8* (4), 297–315. 52

De Loecker, Jan, "Detecting learning by exporting," *American Economic Journal: Microeconomics*, 2013, *5* (3), 1–21. 6

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "The Role of Entrepreneurship in US Job Creation and Economic Dynamism," *Journal of Economic Perspectives*, 2014, *28* (3), 3–24. 52

Dhingra, Swati and John Morrow, "The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition," 2012. 21, 24, 36

Disney, Richard, Jonathan Haskel, Ylva Heden, and Jonathan Haskelt, "Entry, Exit and Establishment Survival in UK Manufacturing," *the Journal of Industrial Economics*, 2003, *51* (1), 91–112. 53

Dixit, Avinash, "Analytical Models in Approximations of Hysteresis," *Review of Economic Studies*, 1991, *58* (1), 141–151. 20

Dixit, Avinash K and Robert S Pindyck, *Investment under Uncertainty*, Princeton, NJ: Princeton University Press, 1994. 20, 61

Doms, Mark E., Ron S. Jarmin, and Shawn D. Klimek, "Information technology investment and firm performance in US retail trade," *Economics of Innovation and New Technology*, 2004, *13* (7), 595–613. 5, 52

Doraszelski, U. and J. Jaumandreu, "R&D and Productivity: Estimating Endogenous Productivity," *The Review of Economic Studies*, feb 2013, *80* (4), 1338–1383. 6

Duffie, Darrell and Kenneth J Singleton, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 1993, *61* (4), 929–952. 31

Dunne, Timothy, Mark J Roberts, Larry Samuelson, and Timothy Dunne, "Patterns of firm entry and exit in U . S . manufacturing industries," *RAND Journal of Economics*, 1988, *19* (4), 495–515. 12, 53

Durham, G.B. and A.R. Gallant, *Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes*, Vol. 20 2002. 31

Eeckhout, Jan and Boyan Jovanovic, "Knowledge Spillovers and Patent Citations," *American Economic Review*, 2002, *92* (5), 1290–1307. 3, 11

Ericson, Richard and Ariel Pakes, "Markov-Perfect Industry Dynamics: Framework for Work Empirical," *Review of Economic Studies*, 1995, *62* (1), 53–82. 3, 6, 8, 10, 36

Evans, David S, "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, 1987, *35* (4), 567–581. 52

Foster, L., J. Haltiwanger, and C. J. Krizan, "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s," *Review of Economics and Statistics*, 2006, *88* (4), 748–758. 2, 5, 32, 33, 34

Foster, Lucia, John C Haltiwanger, Shawn D Klimek, C J Krizan, and Scott Ohlmacher, "The evolution of national retail chains: how we got here," in Emek Basker, ed., *Handbook of the Economics of Retailing and Distribution*, Edward Elgar Publishing, 2016. 5

— , **John Haltiwanger, and C. J. Krizan**, *Aggregate Productivity Growth: Lessons from Microeconomic Evidence* number January 2001. 2, 4, 33, 37, 53

Gabaix, Xavier, "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 2011, *79* (3), 733–772. 54

— , "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 2016, *30* (1), 185–206. 33

Garcia-Macia, Daniel, Hsieh, and Peter J Klenow, "How Destructive is Innovation?," 2015. 7

Geerolf, François, "A Static and Microfounded Theory of Zipf' s Law for Firms and of the Top Labor Income Distribution," 2014. 52

Gibrat, Robert, *Les inégalités économiques*, Paris: Recueil Sirey, 1931. 52

Gourieroux, Christian and Alain Monfort, "Simulation-based inference. A survey with special reference to panel data models," *Journal of Econometrics*, 1993, *59* (1-2), 5–33. 31

Griliches, Zvi, "Issues in assessing the contribution of research and development to productivity growth," *The Bell Journal of Economics*, 1979, *10* (1), 92–116. 6

— , *R&D, Education, and Productivity: A Retrospective*, Harvard University Press, 2000. 52

Grossman, Gene M and Elhanan Helpman, "Trade, innovation, and growth," *American Economic Review*, feb 1990, *151* (3712), 867–8. 2

— **and** — , *Innovation and Growth in the Global Economy*, Massachusetts Intitute of Technology, 1991. 21

Hall, Bronwyn and John Van Reenen, "How effective are fiscal incentives for R&D? A review of the evidence," *Research Policy*, apr 2000, *29* (4-5), 449–469. 30

Hall, Bronwyn H, "The relationship between firm size and firm growth in the US manufacturing sector," *Journal of Industrial Economics*, 1987, *34* (2), 217–223. 52

— , "R&D tax policy during the 1980s: success or failure?," in James Poterba, ed., *Tax Policy and the Economy, Vol 7*, Vol. 7, MIT Press, 1993, pp. 1–36. 30

— , "The private and social returns to research and development," in Bruce L R Smith and Claude E Barfield, eds., *Technology, R&D, and the Economy*, Brookings Institute, 1996. 52

Haltiwanger, John, "Job Creation, Job Destruction, and Productivity Growth: The Role of Young Businesses," *Annual Review of Economics*, 2014, *7* (1), 150504160918000. 52

Haltiwanger, John C, Ron S Jarmin, and Javier Miranda, "Who creates jobs? Small vs. large vs. young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361. 52

Haltiwanger, John, Ron Jarmin, and Javier Miranda, "Business Dynamics Statistics: An Overview," Technical Report January 2009. 29

Hart, P E and S J Prais, "The analysis of business concentration: a statistical approach," *Journal of the Royal Statistical Society. Series A (General)*, 1956, *53* (2), 365–375. 52

Helpman, Elhanan, Marc J Melitz, and Stephen R Yeaple, "Export versus FDI with Heterogeneous Firms," *American Economic Review*, 2004, *94* (1), 300–316. 52

Hermalin, Benjamin E, "Leadership and corporate culture," in Robert Gibbons and John Roberts, eds., *Handbook of Organizational Economics*, Princeton University Press, 2012, pp. 432—-478. 10

Holmes, Thomas J, "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 2011, *79* (1), 253–302. 32

Hopenhayn, Hugo A, "Entry, Exit, and firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992, *60* (5), 1127–1150. 6, 36

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 14031–448. 30

Huber, Peter, Harald Oberhofer, and Michael Pfaffermayr, "Who Creates Jobs ? Estimating Job Creation Rates at the Firm Level," *WIFO working paper*, 2013. 52

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670. 6

— **and Glenn M Macdonald**, "Competitive Diffusion," *Journal of Political Economy*, 1994, *102* (1), 24–52. 6

— **and Glenn M MacDonald**, "Competitive diffusion," *Journal of Political Economy*, 1994, *102* (3), 547–565. 52

— **and** — , "The Life Cycle of a Competitive Industry," *Journal of Political Economy*, 1994, *102* (2), 322–347. 36

— **and Rafael Rob**, "The Growth and Diffusion of Knowledge," *The Review of Economic Studies*, 1989, *56* (4), 569–582. 6

Klepper, Steven, "Entry, exit, growth, and innovation over the product life cycle," *American Economic Review*, 1996, *86* (3), 562–583. 36

— **and Elizabeth Graddy**, "The evolution of new industries and the determinants of market structure," *The RAND Journal of Economics*, 1990, *21* (1), 27–44. 36

— **and Kenneth L. Simons**, "The Making of an Oligopoly: Firm Survival and Technological Change in the Evolution of the U.S. Tire Industry," *Journal of Political Economy*, 2000, *108* (4), 728–760. 36

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018. 2, 6, 7, 9, 10, 15, 33, 52, 53

König, Michael D., Jan Lorenz, and Fabrizio Zilibotti, "Innovation vs imitation and the evolution of productivity distributions," 2015, *11*, 1–48. 6

Lentz, Rasmus and Dale T. Mortensen, "An Empirical Model of Growth Through Product Innovation," *Econometrica*, 2008, *76* (6), 1317–1373. 31, 33, 52, 53

— **and Dale T Mortensen**, "Optimal Growth Through Product Innovation," 2014. 2, 6

Lucas, Robert E and Benjamin Moll, "Knowledge growth and the allocation of time," *Journal of Political Economy*, 2014, *122* (1). 6, 14

Luttmer, Erzo G J, "On the mechanics of firm growth," *Review of Economic Studies*, 2011, *78* (3), 1042–1068. 52

— , "Technology diffusion and growth," *Journal of Economic Theory*, 2012, *147* (2), 602–622. 3, 6, 11, 12, 13, 17

Luttmer, Erzo G.J., "Selection, growth, and the size distribution of firms," *Quarterly Journal of Economics*, 2007, *122* (3), 1103–1144. 3, 6, 11, 12, 52

— , "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, sep 2010, *2* (1), 547–576. 4, 13, 14, 52

McFadden, Daniel, "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 1989, *57* (5), 995–1026. 31

McKinsey Global Institute, "U.S. Productivity Growth 1995-2000," Technical Report, McKinsey Global Institute, Washington, DC 2001. 4

Melitz, Marc J, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725. 3, 24, 36

Mizuno, T, A Ishikawa, S Fujimoto, and T Watanabe, "Power laws in firm productivity," *Progress of Theoretical Physics Supplement*, 2012, (194), 122–134. 52

Neumark, David, Brandon Wall, and Junfu Zhang, "Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series," *Review of Economics and Statistics*, 2011, *93* (1), 16–29. 52

OECD, "R&D tax incentives: rationale, design, evaluation," Technical Report, Organization for Economic Cooperation and Development Innovation Platform 2010. 2

— , "International Experience with R&D Tax Incentives," 2011. 2

— , "OECD Science, Technology and Industry Outlook 2012," Technical Report 2012. 2

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–1297. 6, 52

Pakes, Ariel and David Pollard, "Simulation and the asymptotics of optimization estimators," *Econometrica*, 1989, *57* (5), 1027–1057. 31

Pedersen, Asger R, "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 1995, *22* (1), 55–71. 31

Perla, Jesse and Christopher Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, 2014, *122* (1), 1–21. 3, 6

— , — , and Michael E. Waugh, "Equilibrium Technology Diffusion, Trade, and Growth," 2015. 6, 12, 30

Peters, Michael, "Heterogeneous Mark-Ups and Endogenous Misallocation," 2011. 7

Prescott, Edward C and Michael Visscher, "Organization Capital," *Journal of Political Economy*, 1980, *88* (3), 446–461. 10

Restuccia, Diego and Richard Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, oct 2008, *11* (4), 707–720. 6

Sampson, Thomas, "Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth," *Quarterly Journal of Economics*, 2016, (January). 3, 6, 11, 12

Simon, Herbert A and Charles P Bonini, "The size distribution of business firms," *American Economic Review*, 1958, *48* (1), 1–17. 52

Syverson, Chad, "What Determines Productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365. 2, 10

Wilson, Daniel J, "Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits," *Review of Economics and Statistics*, 2009, *91* (May), 431–436. 30

World Bank, "Innovation policy: A guide for developing countries," Technical Report, World Bank Group 2010. 2

Xu, Daniel Yi, "A Structural Empirical Model of R&D, Firm Heterogeneity, and Industry Evolution," *Working Paper*, 2008, (January), 1–40. 6

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse, "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," *Econometrica*, 2012, *80* (6), 2765–2784. 36

Tables

Targeted Moment		Retail Trade		Manufacturing	
		Model	**Data**	**Model**	**Data**
Rate of productivity growth	g_θ	0.032	0.032	0.038	0.038
Rate of emp growth	g_N	0.017	0.017	0.000	-0.012
Cost elasticity of investment		1.0	1.0	1.0	1.0
Rate of entry		0.117	0.119	0.079	0.079
Rate of emp growth of old (15-20 yr)		-0.015	-0.007	-0.024	-0.017
Rate of reallocation of old (15-20 yr)		0.26	0.23	0.20	0.20
Rate of exit of mature (10-15 yr)		0.075	0.083	0.060	0.060
Rate of exit of old (15-20 yr)		0.065	0.071	0.054	0.052
Parameter					
Volatility	σ	0.19		0.15	
Innovation costs	χ	13360		8103	
Curvature of costs	κ	2.2		2.3	
Entry costs	ψ	6.2		11.3	
Directedness of adoption	μ	3		1	
Decreasing returns of adoption	η	0.6		0.3	
Entry Productivity Shifter	γ_e	-0.5		-0.5	

Table 1: Targeted moments and the calibrated parameters for retail trade and manufacturing. The first three targeted moments each directly pin down a corresponding parameter in the model: the rate of productivity growth pins down the relative cost of entry ψ, the rate of employment growth pins down the rate of growth in the number of firms g_M, and the cost elasticity pins down the curvature parameter of innovation costs κ. The 5 moments of the firm life-cycle dynamics together calibrate the five remaining parameters of the model, characterizing the volatility of firm outcomes σ, the costs of innovation χ, and the process of entry and adoption (μ, η, γ_e).

Variable	Retail Trade			Manufacturing		
	Before	**After**	**Change**	**Before**	**After**	**Change**
Rate of Productivity Growth (%)	3.22	3.51	0.28	3.76	4.13	0.37
... contribution of Turnover	0.774	0.772	-0.002	0.538	0.547	0.009
... contribution of Net Entry	-0.776	-0.775	0.001	-0.005	-0.005	0.000
Rate of Entry	0.118	0.119	0.001	0.079	0.081	0.002
Entry/Exit Productivity Gap	0.519	0.524	0.005	0.795	0.798	0.004
Concentration (Pareto Index)	1.167	1.161	-0.006	1.199	1.196	-0.003
Dispersion of Investments (IDR)	0.377	0.395	0.018	0.361	0.380	0.019

Table 2: The impact of a drastic fall in the costs of innovation (parameter χ) to 75% of its initial value on industry outcomes. For each sector, the leftmost column, under the heading "Before," shows the value of the variable under the calibrated parameters, the middle column, under the heading "After," shows the value after the change. The third column then presents the difference between the two values. IDR stands for Inter-Decile Range.

Variable		Retail Trade	Manufacturing
Optimal Rate of Growth	$g_{\theta,s}$	0.052	0.047
Average Incumbent Subsidy	$-\mathbb{E}_H[\tau_{i,s}(z)]$	0.62	0.28
Entrant Subsidy	$-\tau_{e,s}$	0.45	0.18
Operation Tax	$\tau_{o,s}$	4.90	2.95
Baseline Subsidy Rate		0.62	0.23
Slope of the Subsidy Rate		0.02	0.10
Optimal Schedule Spending		0.32	0.20
Best Uniform Spending		2.57	1.96

Table 3: The characteristics of the optimal allocations of innovation and production, as well as a subsidy scheme that implements them, in retail trade and manufacturing. See Section 4.5 for details.

Figures

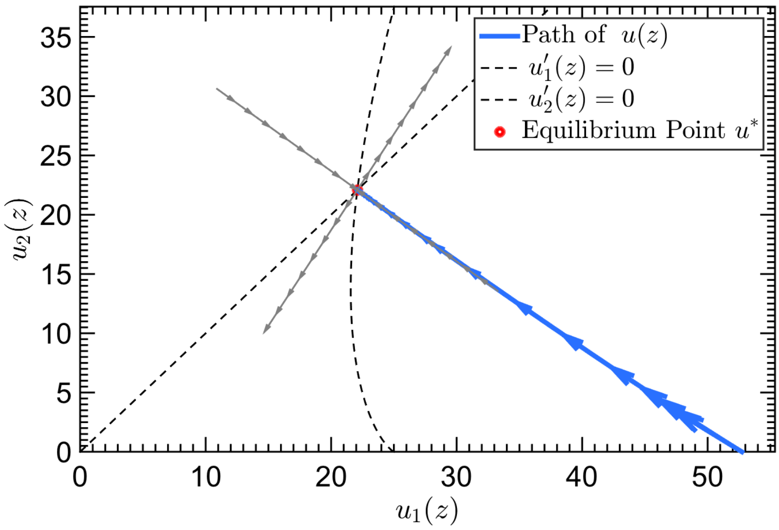

Figure 1: The solution to the Bellman equation in the two-dimensional space of $(u_1(z), u_2(z))$ for a given choice of model parameters (details in the online appendix). The relative sizes of arrows on the path of $u(z)$ indicate the relative values of the derivative of u with respect to z along the path. The equilibrium point u^*, which lies on the 45^o line, is shown with a red dot. The stable arm around the saddle-path stable equilibrium point is also shown in light gray. For very large firms, as $z \to \infty$, the path of $u(z)$ asymptotically converges to this arm.

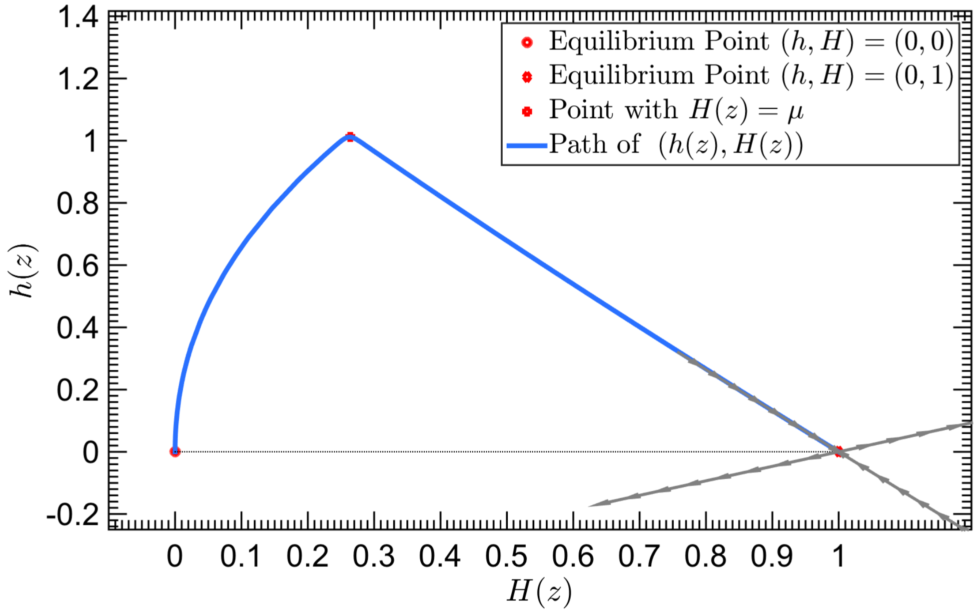

Figure 2: The solution to the differential equation characterizing the stationary distribution in the space of $(h(z), H(z))$ for a given choice of model parameters and an atomic distribution of adoption F_a (details in the online appendix). The two equilibrium points and the point of entry are shown in red. The stable arm around the saddle-path stable equilibrium point $(H, h) = (1, 0)$ is shown in light gray. For very large firms, as $z \to \infty$, the path of (H, h) asymptotically converges to the stable arm.

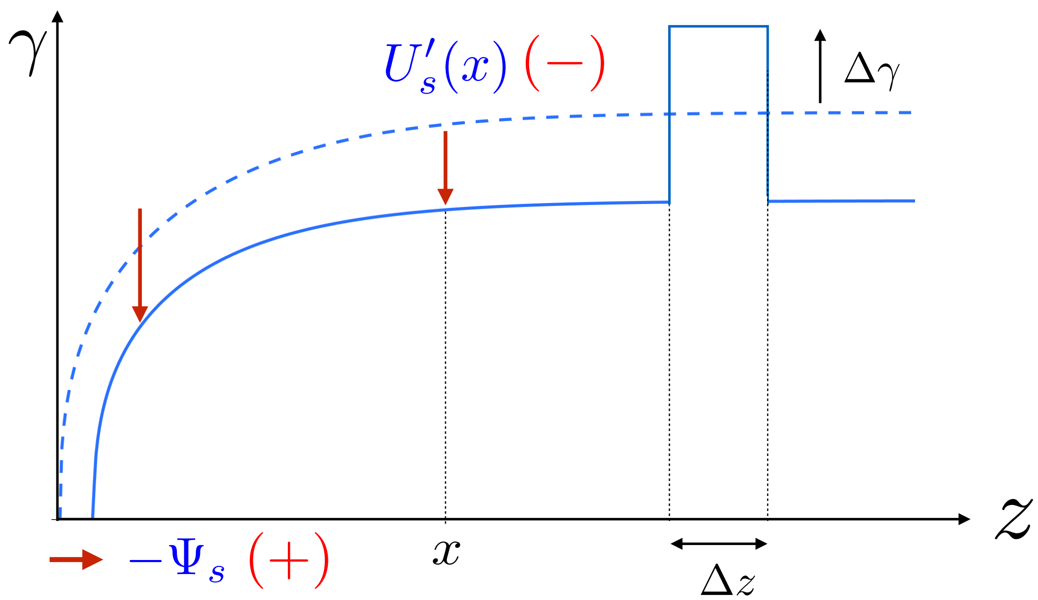

Figure 3: The illustration of the dynamic displacement externalities of firm innovation using a partial equilibrium perturbation exercise. Starting from the dashed, market-equilibrium schedule of innovation investments as a function of efficiency states, we push the investments of firms in the band Δz aroudn z upward by $\Delta\gamma$ to find the solid schedule. For firms at efficiency state x, the negative impact is proportional to $U_s'(x)$. The effect for firms near the exit threshold is positive.

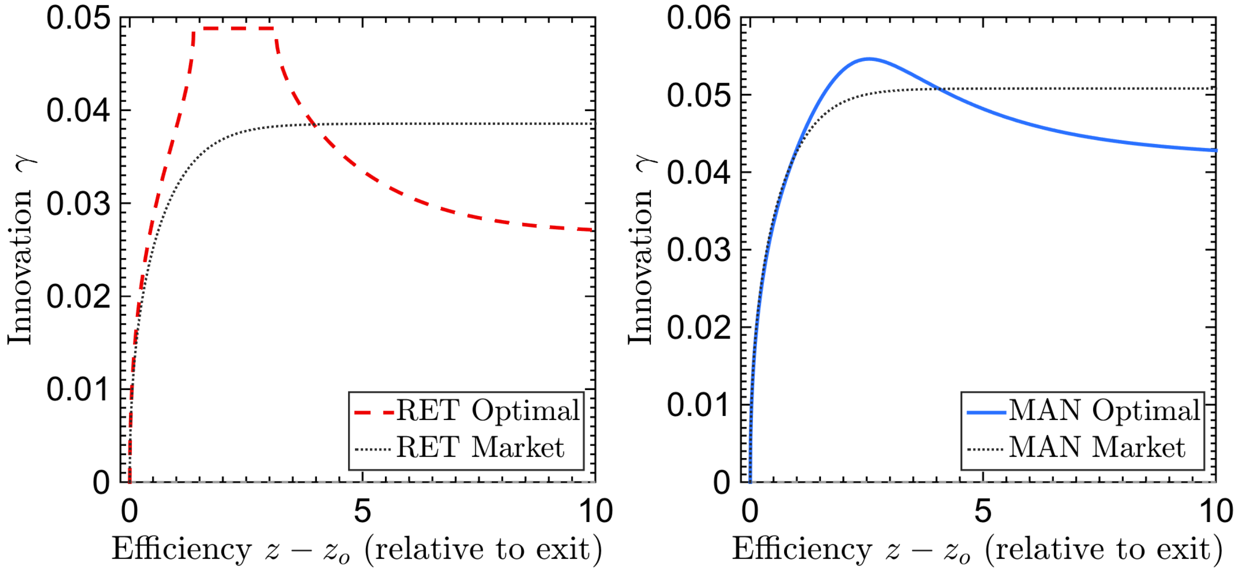

Figure 4: Comparison between the optimal and market equilibrium allocations of innovation in (Right) manufacturing and (Left) retail, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.

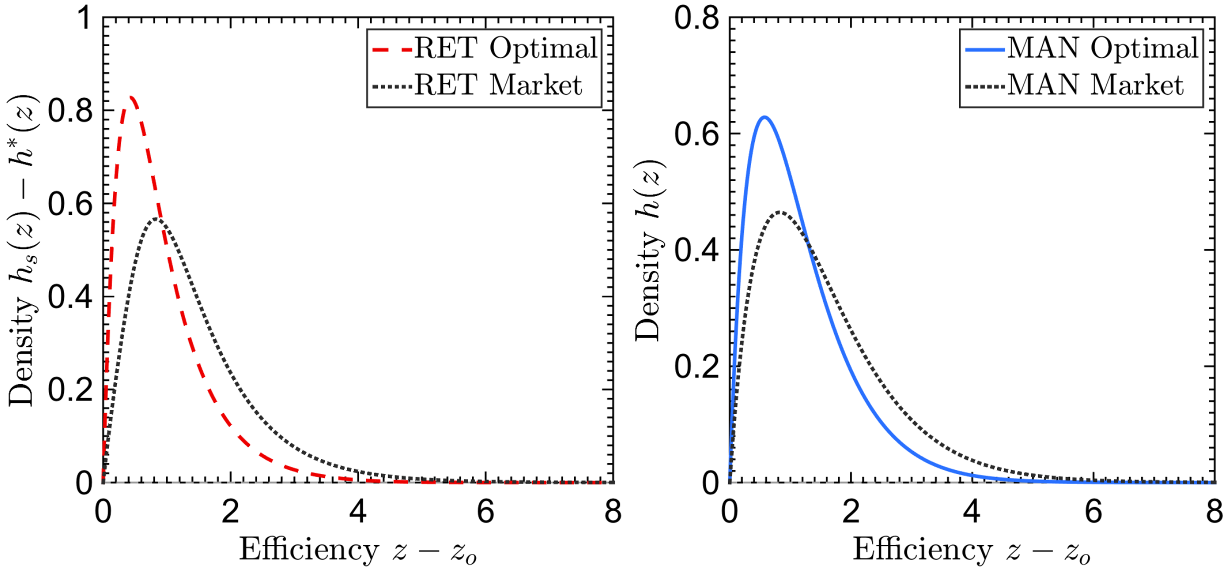

Figure 5: Comparison between the optimal and market equilibrium long-run distributions of efficiency in (Right) manufacturing and (Left) retail. RET and MAN stand for retail and manufacturing, respectively.

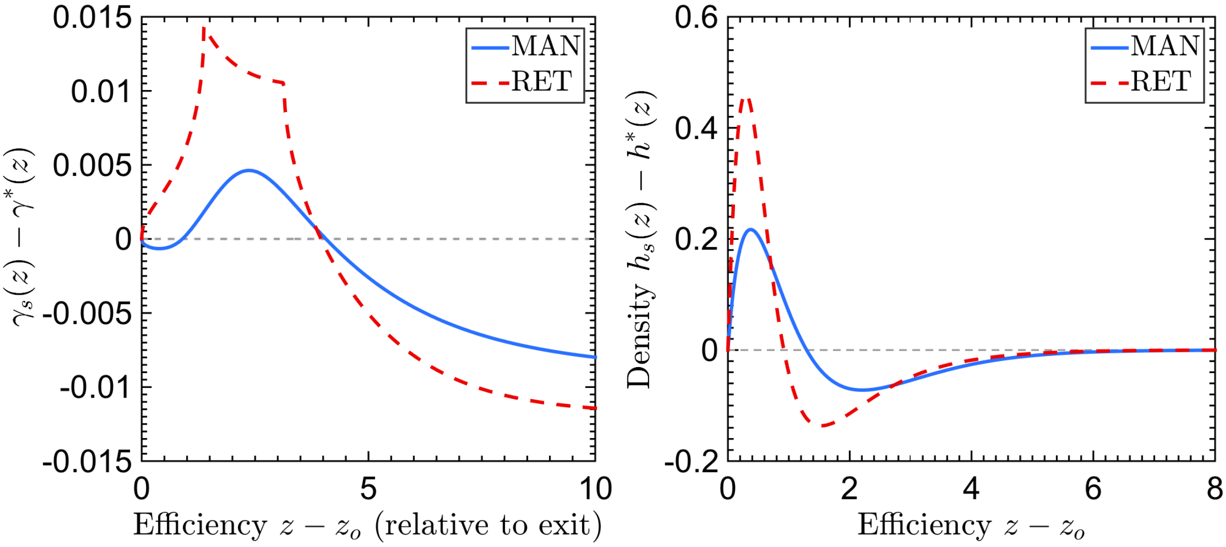

Figure 6: The difference between the optimal and market equilibrium allocations in (Right) innovation investments and (Left) long-run probability distribution function, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.

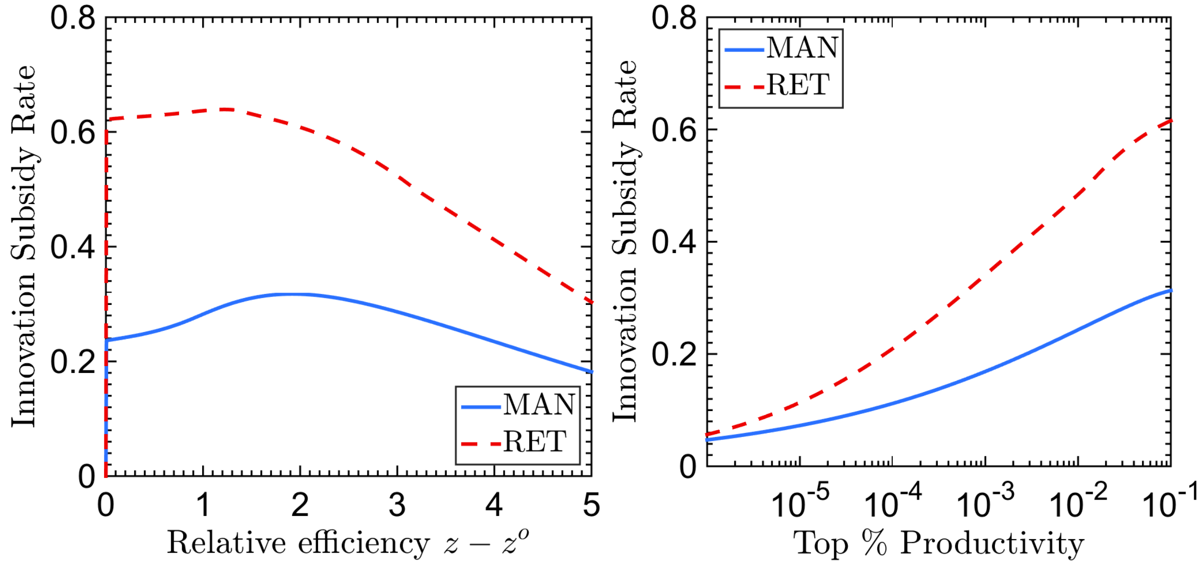

Figure 7: (Left) Comparison between the subsidy scheme that achieves the optimal innovation investments in manufacturing and retail, as a function of the relative efficiency of firms. (Right) The same comparison but zooming in on the right tail of the productivity distribution. Here, the x-axis shows the percentile ranking of the productivity of firms at the top of the distribution of productivities. RET and MAN stand for retail and manufacturing, respectively.

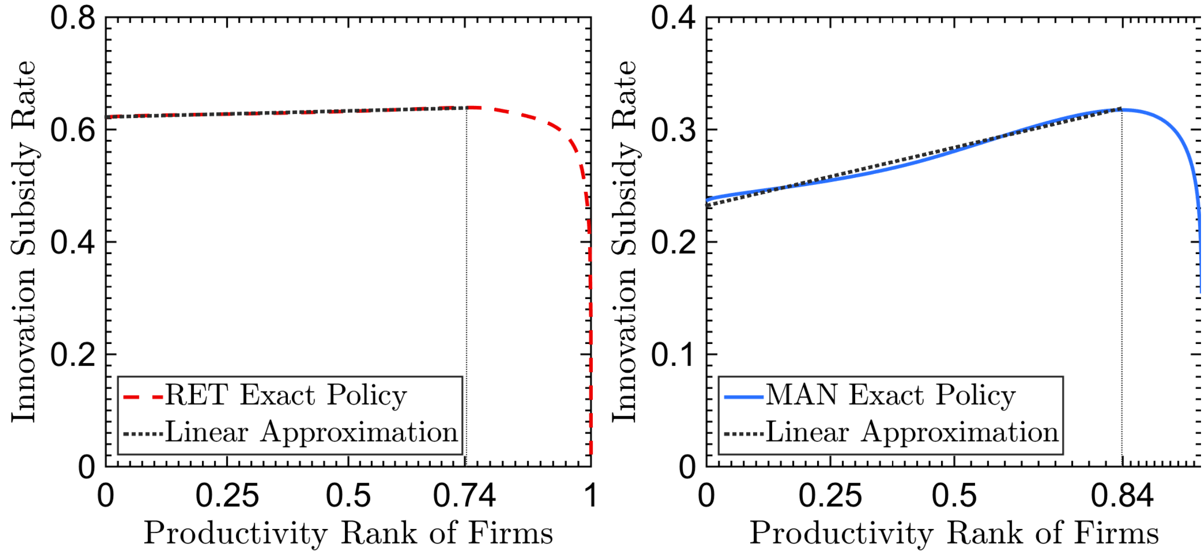

Figure 8: Comparison between the exact and the linearly approximated optimal subsidy for (Left) retail and (Right) manufacturing. The gray vertical line indicates the rank corresponding to the maximum of the subsidy rate.

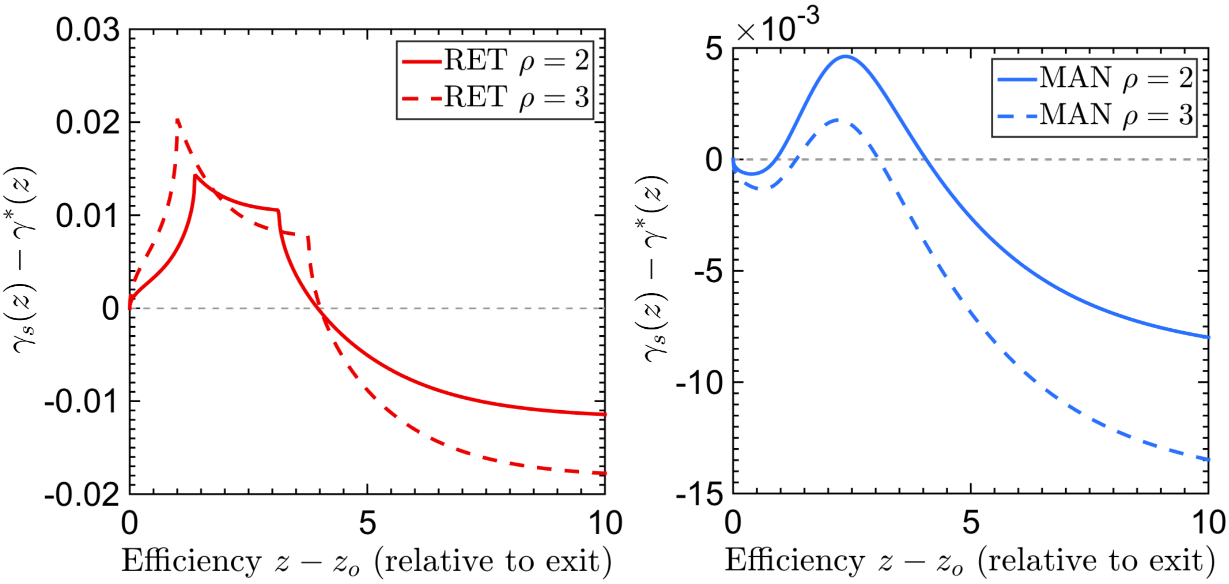

Figure 9: Comparison of the difference between the optimal and market equilibrium allocations with two different degrees of substitutability in firm outputs in (Right) retail and (Left) manufacturing, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.

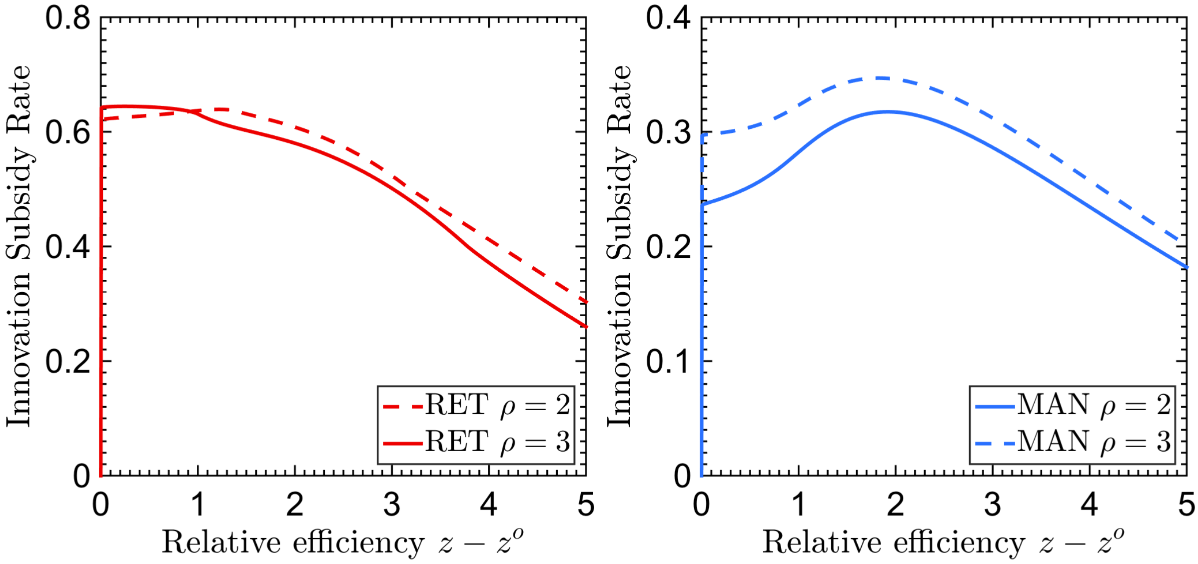

Figure 10: Comparison between the subsidy scheme that achieves the optimal innovation investments with two different degrees of substitutability in firm outputs in (Left) retail and (Right) manufacturing, as a function of the relative efficiency of firms. RET and MAN stand for retail and manufacturing, respectively.

6 Facts

In this section, I provide an overview of the set of facts on the distribution and dynamics of firm size and productivity that the model aims to explain. I also discuss how the existing theories of firm dynamics generate these facts.

1. Firm size and productivity distributions exhibit Pareto tail behavior In the US data, the logarithm of the complementary cumulative distribution function (tail distribution) of firm size is linear in the logarithm of firm size in its right tail (Axtell, 2001; Luttmer, 2010).[47] There is also evidence that the distribution of total factor productivity of firms across different countries has a Pareto tail (Aoyama et al., 2010; Mizuno et al., 2012).[48] Among the theories of firm heterogeneity and dynamics that endogenize the firm size distribution, the models of Luttmer (2007), Luttmer (2011), Arkolakis (2016), and the current paper match with this fact. In contrast, the benchmark model of Klette and Kortum (2004) and its empirical extension by Lentz and Mortensen (2008) both yield considerably thinner right tails. Luttmer (2010) gives an overview of different approaches and necessary assumptions for generating a Pareto tail of firm size.

2. Size and productivity are correlated Following Olley and Pakes (1996), a large body of empirical work has utilized the covariance of size and productivity as a measure of aggregate industry productivity (see Bartelsman et al., 2013, for cross-country evidence). Under the baseline of the workhorse innovation model of Klette and Kortum (2004), firm size and productivity are independent of each other due to the undirectedness of the innovation process. Therefore, we need to introduce further heterogeneity among firms in terms of the quality of their innovations (e.g., Lentz and Mortensen, 2008). In contrast, the productivity-based class of models readily explains the correlation between size/productivity, since the variations in size derive from productivity heterogeneity in the presence of decreasing returns to scale or imperfect competition.

3. Growth is negatively correlated with size for small firms and uncorrelated with size for large firms Based on early firm-level data, Gibrat and subsequent researchers found that firm growth does not depend on size (e.g., Hart and Prais, 1956; Simon and Bonini, 1958). Using US manufacturing data, Hall (1987) and Evans (1987) revisited Gibrat's law accounting for the role of selection and measurement error, and concluded that the relationship is negative for small firms, but approximately holds for larger firms.[49] Models of firm dynamics match relying on different mechanisms to explain this fact. In the current paper, the faster growth of smaller firms is driven by stronger selection.[50] Other recent papers such as Akcigit and Kerr (2012) and Acemoglu et al. (2014) deliver this prediction by introducing further heterogeneity in the types and quality of innovations across firms.

4. The extensive margin of innovation intensity is correlated with productivity and size The evidence on this point mainly comes from the early studies of R&D, size, and productivity (see Griliches, 2000; Hall, 1996, for surveys of this literature). The earlier evidence based on manufacturing firms generally emphasized the correlation between the extensive margin of innovation investments (e.g., whether or not firms report any R&D expenditures) and productivity or size (e.g., Cohen and Klepper, 1996). In the US retail sector, Doms et al. (2004) find a pronounced correlation between firm size and productivity and intensive measures of investment in IT. In the presence of very small fixed costs of innovation, the model in this paper delivers this fact through the linkage between selection and innovation incentives. This is because larger and more productive firms have higher incentives for innovation as they have a longer

[47] Geerolf (2014) provides evidence for the Pareto tail in the distribution of firm size in France and constructs an alternative, static, matching-based model that can rationalize this phenomenon.

[48] This is indeed a standard assumption in many models of firm heterogeneity and trade (see, e.g., Helpman et al., 2004; Chaney, 2008).

[49] Davis et al. (1996) used US census data to make the case that controlling for age, this relationship does not hold even among smaller firms (see also Haltiwanger et al., 2013; Decker et al., 2014; Haltiwanger, 2014). However, debate around this point still continues, as some recent studies challenge the conclusions of Davis et al. (1996) based on other datasets and different empirical specifications (Neumark et al., 2011; Huber et al., 2013). See also the discussions about the details of these specifications in (Akcigit and Kerr, 2012).

[50] An example of an alternative approach is the model by Jovanovic and MacDonald (1994b), in which all firms look within the same pool of available ideas. Firms that are ahead have further exhausted the space of potential new ideas and have less to gain from their investments.

expected lifetime in the industry. However, focusing on R&D expenditure and patents as measures of innovation investments and outputs, respectively, Akcigit and Kerr (2012) find a negative relationship between innovation intensity and size among a subsample of manufacturing firms that they refer to as "innovative firms," i.e., those that do report R&D expenditures over an extended period of time and also file for patents. In explaining the extensive margin of innovation and considering a broader concept of innovation activity, the current paper could be complementary to more refined theories of firm-level R&D and patenting behavior (Akcigit and Kerr, 2012).

5. Likelihood of exit is negatively correlated with productivity and size The negative relationship between size and likelihood of exit is nearly universal. Foster et al. (2001) summarize the evidence that shows that low productivity predicts exit even controlling for a variety of different firm characteristics. The current model explains this fact through the assumption on the fixed costs, which gives rise to the endogenous exit of low-productivity firms. In the workhorse model of Klette and Kortum (2004), since size and productivity are independent we do not necessarily find a correlation between productivity and size. Once again, introducing additional heterogeneity in terms of innovation quality within this framework can produce the correlation between size and productivity in the cross section (e.g., Lentz and Mortensen, 2008). More recent work has combined competition through the input markets (as in the current paper) with the heterogeneity in innovative types within the workhorse framework (Acemoglu et al., 2013).

6. Entrants are smaller, less productive, and less dispersed than incumbents The evidence is widespread in studies of comprehensive firm-level data from industries across the world. Seminal studies reporting the evidence include, among others, Dunne et al. (1988) and Disney et al. (2003).

7 Derivations and Proofs

7.1 Section 2

Dynamic Industry Equilibrium Let $v_t(\theta)$ denote the value of a firm with productivity θ at time t. Due to the presence of fixed costs, firms endogenously exit the industry once their productivity falls below an exit threshold. Let $\theta_{o,t}$ denote the exit threshold of incumbent firms at time t. The value function then satisfies a Hamilton-Jacobi-Bellman (HJB) partial differential equation for all productivity states $\theta > \theta_{o,t}$ above this threshold (see Equation (44) below). The solution to the HJB equation yields the value function $[v_t(\cdot)]_t$ as well as an incumbent firm's path of optimal innovation and exit policies $\varsigma^* \equiv [\gamma_t^*(\cdot), \theta_{o,t}]_{t=0}^{\infty}$.

Define the cumulative distribution of entrant productivity (before entry) at time t as

$$F_{e,t}^{Pre}(\theta) \equiv F_a\left[M_t\left(\left(e^{-\gamma_e}\theta_{o,t}\tilde{\theta}_t{}^{\eta}\right)^{1/(1-\eta)}\right)/\overline{M}_t\right].$$

Let $\lambda_{e,t} \geq 0$ denote the rate of entry at time t, defined as the flow of entrants divided by the total mass of currently active firms in the industry. Standard arguments show that along any equilibrium path, the following zero profit condition holds

$$c_{e,t} \geq v_{e,t}, \text{ and } \lambda_{e,t}(c_{e,t} - v_{e,t}) = 0, \qquad \forall t \geq 0, \tag{42}$$

where the costs of entry are given by $c_{e,t} \equiv \psi_e w_t$ and the expected net presented value $v_{e,t}$ by

$$v_{e,t} \quad \equiv \quad \int v_t(\theta)\, dF_{e,t}(\theta).$$

Now consider such an optimal policy ς^* and define the mass density of firms as $m_t(\theta) = \frac{\partial}{\partial\theta}M_t(\theta)$. We can show that the industry measure has to satisfy a Kolmogorov Forward Equation (KFE) (see Equation (46) in Appendix 7.1). Along with boundary conditions $M_t(\theta_{o,t}) = 0$ and $\overline{M}_t < \infty$ (for all t), this partial differential equation characterizes the path of industry measure $[M_t(\cdot)]_t$ and the rate of turnover (exit) $\lambda_{o,t}$ satisfying

$$\overline{M}_t\lambda_{o,t} \equiv \frac{1}{2}\sigma^2\frac{\partial}{\partial\theta}\left(\theta^2 m_t(\theta)\right)\bigg|_{\theta=\theta_{o,t}}, \tag{43}$$

for a given initial industry measure M_o. Note that the rate of growth of the total mass of firms in the industry is given by the difference between rate of (successful) entry $\lambda_{e,t}$ and exit $\lambda_{o,t}$, such that $\dot{\overline{M}}_t/\overline{M}_t = \lambda_{e,t} - \lambda_{o,t}$. If $\gamma_e > -\eta \log\left(\tilde{\theta}_t/\theta_{o,t}\right)$, then entrants do not exit immediately, the gross rate of entry equals the successful rate of entry $\lambda_{o,t}$, and the post-entry distribution of entrants is the same as the pre-entry one: $F_{e,t}(\theta) \equiv F^{Pre}_{e,t}(\theta)$. Otherwise, the gross rate of entry is greater than the rate of successful entry by a factor of $\left[1 - F^{Pre}_{e,t}(\theta_{o,t})\right]^{-1}$. In this case, the post-entry distribution of entrants is given by

$$F_{e,t}(\theta) \equiv \frac{F^{Pre}_{e,t}(\theta) - F^{Pre}_{e,t}(\theta_{o,t})}{1 - F^{Pre}_{e,t}(\theta_{o,t})}.$$

In the dynamic industry equilibrium, firms take the future paths of wages, interest rate, and aggregate sales as given, rationally expect the behavior of other firms, and decide on their optimal strategies.[51] The aggregation of all these decisions generates the path of the industry measure $[M_t(\cdot)]_{t=0}^{\infty}$ that contains all the information regarding the state of the industry.

Definition 2. (Industry Equilibrium) Given an initial industry measure M_0, a time path of wages $[w_t]_{t=0}^{\infty}$, interest rate r, and a time path of industry aggregate sales $[N_t]_{t=0}^{\infty}$, an industry tuple $\boldsymbol{\mathcal{I}} \equiv \left[M_t(\cdot), \gamma_t^*(\cdot), \theta_{o,t}, \lambda_{e,t}\right]_{t=0}^{\infty}$ characterizes a dynamic industry equilibrium if

1. The policy function $\varsigma \equiv \left[\gamma_t^*(\cdot), \theta_{o,t}\right]_{t=0}^{\infty}$ constitutes a solution to the HJB equation (44) (see the derivation below),
2. Rate of entry $\lambda_{e,t}$ satisfies (42) for the corresponding value function,
3. The measure $M_t(\cdot)$ satisfies the KFE (46) (see the derivation below) with the corresponding boundary conditions.

General Equilibrium I introduce the market clearing conditions and use Definition 2 for the industry equilibrium to define the general equilibrium of this economy.

Definition 3. (General Equilibrium) Consider an industry tuple $\boldsymbol{\mathcal{I}}$ and a path of population $[N_t]_{t=0}^{\infty}$ such that

$$\frac{N_t}{\overline{M}_t} > \psi_f\left(1 + \psi\lambda_{e,t}\right),$$

for all t. A path of wages, consumption, and per-capita assets $[w_t, q_t, a_t]_{t=0}^{\infty}$ and the industry $\boldsymbol{\mathcal{I}}$ together form a dynamic general equilibrium if we have

1. Given the path of wages $[w_t]_{t=0}^{\infty}$, interest rate $r_t = r$, and aggregate sales (and population) N_t, the industry constitutes a dynamic industry equilibrium,
2. The paths of wages, per-capita consumption, and per-capita assets satisfy

$$\begin{aligned}
w_t &= \frac{\rho}{\rho+1}\frac{N_t}{N_t - \overline{M}_t\psi_f\left(1+\psi\lambda_{e,t}\right)}, \\
q_t &= \frac{\rho}{\rho+1}\frac{\overline{\theta}_t}{w_t}, \\
a_t &= \frac{1}{N_t}\int dM_t(\theta)\,\left[v_t(\theta) + \psi_e w_t \lambda_{e,t}\right],
\end{aligned}$$

   for all t, where $\overline{\theta}_t$ satisfies (7), and a_t defined above satisfies the transversality condition.

Derivation of the HJB and the KFE Equations The value function for a firm with productivity θ at time t has to satisfy the Bellman equation

$$r\,v_t(\theta)\,dt = \max_{\varsigma}\left[\pi_t(\theta,\gamma;M_t) - \psi_f w_t\right]dt + \mathbb{E}_t\left[dv_t(\theta)\right],$$

[51] Due to the assumption of a continuum of firms, there is no aggregate uncertainty in this model. If we allow for a finite number of firms, then aggregate uncertainty emerges, following the arguments presented by Gabaix (2011) and Carvalho and Grassi (2015).

where π_t denotes the instantaneous profit function as in Equation (4) and $\varsigma \equiv (\gamma_t(\cdot), \theta_{o,t})$ is a tuple of firm policy choice. Using Itô's lemma, we find

$$r\, v_t(\theta) - \frac{\partial v_t(\theta)}{\partial t} = \max_{\gamma, \theta_o} \pi_t(\theta, \gamma) - \psi_f w_t + \gamma_\theta \theta \frac{\partial v_t(\theta)}{\partial \theta} + \frac{\sigma^2}{2}\theta^2 \frac{\partial^2 v_t(\theta)}{\partial \theta^2}. \tag{44}$$

In addition, we have the standard value matching and smooth pasting boundary conditions $v_t(\theta_{o,t}) = \frac{\partial}{\partial \theta} v_t(\theta_{o,t})$ at the exit threshold $\theta_{o,t}$.

Next, I derive the evolution of the industry measure. Let $M_t^b(\theta)$ denote the mass of firms with profitability states below θ at time t that belong to cohort b, and $m_t^b(\cdot)$ denote the corresponding density mass function. For a given choice of optimal firm policies $\varsigma = (\gamma_t(\cdot), \theta_{o,t})$, the KFE characterizes the evolution of the measure over time as

$$\frac{\partial m_t^b(\theta)}{\partial t} = -\frac{\partial}{\partial \theta}\left[\gamma_t^*(\theta)\, \theta\, m_t^b(\theta)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2 \theta}\left[\theta^2 m_t^b(\theta)\right],$$

subject to the boundary conditions $\lim_{t \downarrow b} M_t^b(\theta) = M_t^t(\theta) = \overline{M}_t \lambda_{e,t} F_{e,t}(\theta)$ and $m_t^b(\theta_{o,t}) = 0$, $\forall t, \forall b,\ t > b$.

We can now integrate the mass over all cohorts to find the industry-wide distribution of firm efficiencies

$$M_t(\theta) = \int_{-\infty}^{t} M_t^b(\theta)\, db, \qquad m_t(\theta) = \int_{-\infty}^{t} m_t^b(\theta)\, db.$$

Applying the same integration to the KFE above

$$\begin{aligned} \frac{\partial m_t(\theta)}{\partial t} &= m_t^t(\theta) + \int_{-\infty}^{t}\left\{-\frac{\partial}{\partial \theta}\left[\gamma_t^*(\theta)\, \theta\, m_t^b(\theta)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2\theta}\left[\theta^2 m_t^b(\theta)\right]\right\} db, \\ &= \overline{M}_t \lambda_{e,t} f_{e,t}(\theta) - \frac{\partial}{\partial \theta}\left[\gamma_t(\theta)\, \theta m_t(\theta)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2\theta}\left[\theta^2 m_t(\theta)\right], \end{aligned} \tag{45}$$

where I passed the integral through the derivative in the second line. This leads us to the following partial differential equation for the industry measure

$$\frac{\partial M_t(\theta)}{\partial t} = \overline{M}_t\left[\lambda_{e,t} F_{e,t}(\theta) - x_t\right] - \gamma_t^*(\theta)\, \theta m_t(\theta) + \frac{\sigma^2}{2}\frac{\partial}{\partial \theta}\left[\theta^2 m_t(\theta)\right], \tag{46}$$

with boundary conditions $M_t(\theta_{o,t}) = m_t(\theta_{o,t}) = 0$.

To derive the evolution of the total mass of firms, we integrate the equation above to find

$$\dot{M}_t(\theta) = \overline{M}_t \lambda_{e,t} F_{e,t}(\theta) - \left[\gamma_t(\theta)\, \theta m_t(\theta)\right]_{\theta t}^{\theta} + \frac{\sigma^2}{2}\frac{\partial}{\partial \theta}\left[\theta^2 m_t(\theta)\right]_{\theta_{o,t}}^{\theta},$$

which implies

$$\dot{\overline{M}}_t = \overline{M}_t \lambda_{e,t} - \frac{\sigma^2}{2}\frac{\partial^2}{\partial^2 \theta}\left[\theta^2 m_t(\theta)\right]\Big|_{\theta = \theta_{o,t}}.$$

Derivations of the SCGP HJB and KFE Equations I will first show that firm policy functions are stationary functions of z_t along a stationary constant growth path. This will establish that the mapping between firm employment and efficiency z_t remains constant over time. Therefore, the search for a stationary distribution of firm employment corresponds to a search for a stationary distribution of efficiency z_t. I then characterize the necessary conditions that this stationary distribution needs to satisfy.

Let us begin by deriving the value function as a function of the efficiency state. Define function $V_t(z)$ such that

$$v_t(\theta) = \psi_f\, w_t V_t\left(\log\left(\frac{1}{(1+\rho)\,\psi_f w_t}\frac{N_t}{\overline{M}_t}\left(\frac{\theta}{\overline{\theta}_t}\right)^{\rho}\right)\right).$$

Taking the derivative of this equation with respect to time, we find

$$\frac{1}{\psi_f w_t}\frac{\partial v_t(\theta)}{\partial t} = \frac{\dot{w}_t}{w_t} V_t + \frac{\partial V_t}{\partial t} + \left[\frac{\dot{N}_t}{N_t} - \frac{\dot{\overline{M}}_t}{\overline{M}_t} - \frac{\dot{w}_t}{w_t} - \rho\frac{\dot{\overline{\theta}}_t}{\overline{\theta}_t}\right]\frac{\partial V_t}{\partial z}.$$

Similarly, the first and second partial derivative with respect to θ yield:

$$\begin{aligned}
\frac{1}{\psi_f w_t}\frac{\partial v_t(\theta)}{\partial \theta} &= \frac{\rho}{\theta}\frac{\partial V}{\partial z}, \\
\frac{1}{\psi_f w_t}\frac{\partial^2 v_t(\theta)}{\partial^2 \theta} &= \left(\frac{\rho}{\theta}\right)^2\frac{\partial^2 V}{\partial z^2} - \frac{\rho}{\theta^2}\frac{\partial V}{\partial z}.
\end{aligned}$$

Substituting for these expressions in the HJB equation (44) yields:

$$\left(r - \frac{\dot{w}_t}{w_t}\right) V_t(z) - \frac{\partial V_t}{\partial t} = \max_{\gamma} e^z J(\gamma) - 1 + \left[\rho\left(\gamma - \frac{1}{2}\sigma_\theta^2 - \frac{\dot{\bar{\theta}}_t}{\bar{\theta}_t}\right) - \frac{\dot{w}_t}{w_t} + g_N - \frac{\dot{\overline{M}}_t}{\overline{M}_t}\right]\frac{\partial V_t}{\partial z} + \frac{\rho^2}{2}\sigma^2\frac{\partial^2 V_t}{\partial z^2}.$$

Along as a stationary constant growth path, in addition to the conditions in Definition 1, we need the nominal wage rate to converge to a constant w , and hence, $\dot{w}_t/w_t \to 0$. In addition, let us assume for now that $\dot{\overline{M}}_t/\overline{M}_t \to g_N$. We then find

$$r V(z) = \max_{\gamma} e^z J(\gamma) - 1 + \rho\left(\gamma - \frac{1}{2}\sigma^2 - g_\theta\right)\frac{\partial V}{\partial z} + \frac{\rho^2}{2}\sigma^2\frac{\partial^2 V}{\partial z^2},$$

which implies that the optimal investment strategy is also a stationary function of the efficiency state

$$\begin{aligned}
\gamma_t(\theta) &= \gamma^*(z_t(\theta)), \\
z_t(\theta_{o,t}) &= z_o.
\end{aligned}$$

The derivations above suggest that along a stationary constant growth path, we can consider the following form for the industry measure

$$\begin{aligned}
M_t(\theta) &= \overline{M}e^{g_M t} H(z_t(\theta)), \\
m_t(\theta) &= \overline{M}e^{g_M t}\frac{\rho}{\theta} h(z_t(\theta)).
\end{aligned}$$

Once again, we calculate the derivatives:

$$\frac{\partial M_t}{\partial t} = g_M \overline{M}e^{g_M t} H + (g_N - g_M - \rho g_\theta)\overline{M}e^{g_M t} h.$$

and

$$\begin{aligned}
\frac{\partial}{\partial \theta}\left[\theta^2 m_t(\theta)\right] &= \rho\overline{M}e^{g_M t}\frac{\partial}{\partial \theta}\left[\theta h(z_t(\theta))\right], \\
&= \rho\overline{M}e^{g_M t}\left[h(z) + \rho h'(z)\right].
\end{aligned}$$

Substituting these expressions in the Kolmogorov forward equation (46) and dividing both sides by $\overline{M}_t = \overline{M}e^{g_{Mt}}$, we find

$$g_M H = \lambda_e F_e(z) - x - \left(\rho\left(\gamma^*(z) - \frac{1}{2}\sigma^2 - g_\theta\right) + g_N - g_M\right) h(z) + \frac{\rho^2}{2}\sigma^2 h'(z),$$

where $\lambda_o \equiv \frac{\rho^2}{2}\sigma^2 h'_+(z_o)$. If an equilibrium with stationary employment exists, then the industry measure has to be of the form above where H is the solution to this boundary value problem.

Now, we can see why condition $g_M = g_N$ is necessary for the existence of a stationary constant growth path. We have

$$\begin{aligned}
\bar{\theta}_t^\rho &= \frac{1}{\overline{M}_t}\int \theta^\rho J(\gamma_t(\theta))\, dM_t(\theta), \qquad (47)\\
&= \int h(z_t(\theta))\, \theta^\rho J(\gamma^*(z_t(\theta)))\frac{\rho d\theta}{\theta}, \\
&= \left((1+\rho)\, w\, \psi_f \frac{\overline{M}_t}{N_t}\bar{\theta}_t^\rho\right)\int h(z)\, e^z J(\gamma^*(z))\, dz,
\end{aligned}$$

where I have used the fact that $dz = \rho d\theta/\theta$ and $\theta^{\rho} = (1+\rho)\, w\psi_f \frac{\overline{M}_t}{N_t} \overline{\theta}_t^{\rho} e^z$. This implies

$$\frac{N_t}{\overline{M}_t} = (1+\rho)\, w\psi_f \int h(z)\, e^z\, J\left(\gamma^*(z)\right)\, dz.$$

Finally, consider the industry-wide stock of knowledge defined by

$$\tilde{\theta}_t^{\vartheta} = \frac{1}{\overline{M}_t} \int \theta^{\vartheta}\, dM_t(\theta)\,.$$

Along the stationary constant growth path, we can write this as

$$\begin{aligned}
\tilde{\theta}_t^{\vartheta} &= \int h\left(z_t(\theta)\right)\, \theta^{\vartheta}\, \frac{\rho d\theta}{\theta}, \\
&= \left((1+\rho)\, w\, \psi_f \frac{\overline{M}_t}{N_t}\right)^{\frac{\vartheta}{\rho}} \overline{\theta}_t^{\vartheta} \int h(z)\, e^{\vartheta z/\rho} dz.
\end{aligned}$$

We then have

$$\begin{aligned}
e^{z_t(\theta_e)} &= \frac{1}{(1+\rho)\, w\psi_f} \frac{N_t}{\overline{M}_t} \left[\frac{\tilde{\theta}_t^{\eta} \theta_a^{1-\eta} e^{\gamma_e}}{\overline{\theta}_t}\right]^{\rho}, \\
&= \left[\frac{1}{(1+\rho)\, w\psi_f} \frac{N_t}{\overline{M}_t} \left(\frac{\tilde{\theta}_t}{\overline{\theta}_t}\right)^{\rho}\right]^{\eta} \times \left[\frac{1}{(1+\rho)\, w\psi_f} \frac{N_t}{\overline{M}_t} \left(\frac{\theta_a}{\overline{\theta}_t}\right)^{\rho}\right]^{1-\eta} \times e^{\gamma_e}, \\
&= e^{\eta\, \tilde{z} + (1-\eta) z_t(\theta_a) + \rho\gamma_e},
\end{aligned}$$

where, in the last line, I have defined

$$e^{\tilde{z}} \equiv \left(\int h(z)\, e^{\vartheta z/\rho} dz\right)^{\frac{\rho}{\vartheta}}.$$

Proofs of Propositions First, I will provide a generalization of Proposition 1 for the case in which the evolution of efficiencies is given by Equation (15) along with an additional exogenous death that arrives according to a Poisson process with rate δ.

Proposition. *See Proposition 1.*

Proof. I map the problem into a two-dimensional first order differential equation. Use the definitions (27) and (28) and note that

$$\begin{aligned}
V''(z)\, e^{-z} &= u_2'(z) + u_2(z)\,, \\
u_1'(z) &= u_2(z) - u_1(z)\,.
\end{aligned}$$

Now, we can rewrite Equation (17) as

$$(r+\delta)\, u_1(z) = \max_{\gamma} J(\gamma) + \left(\rho\gamma + \frac{\rho^2}{2}\sigma^2 - g\right) u_2(z) + \frac{\rho^2}{2}\sigma^2 u_2'(z)\,,$$

which gives the FOC for the choice of firm growth

$$-J'\left(\gamma^*(z)\right) = u_2(z)\,. \tag{48}$$

Therefore, we can write the differential equation as $\frac{d}{dz}\boldsymbol{u}(z) = \boldsymbol{Q}(\boldsymbol{u}(z), z)$ such that $\boldsymbol{u} \equiv (u_1, u_2)^t$ and $\boldsymbol{Q}$ is given by

$$u_1' = Q_1(\boldsymbol{u}) = u_2 - u_1, \tag{49}$$

$$u_2' = Q_2(\boldsymbol{u}) = \frac{2}{\sigma^2}\left[(r+\delta)\, u_1 - J\left(\gamma^*(z)\right) - \left(\rho\gamma^*(z) - g + \frac{\rho^2}{2}\sigma^2\right) u_2\right], \tag{50}$$

where γ^* is a function of $u_2(z)$ as above. This is a boundary value problem with boundary condition $u_2(z_o) = 0$, which ensures $V'(z_o) = 0$. To find the other boundary conditions, we need to first study the asymptotic behavior of the problem and the existence of a stable equilibrium point for this system of differential equations.

What are the necessary conditions for the existence of an economically meaningful solution to the problem? Let us examine the behavior of the largest firms in the industry. Since the likelihood of exit and the option value of exit shrink to zero for the largest firms, we expect the value function to be linear in e^z for these firms. For a solution satisfying these limiting conditions, we need to ensuare that

$$\lim_{z\to\infty} u_1(z) = \lim_{z\to\infty} u_2(z) = u^*. \tag{51}$$

This implies $\lim_{z\to\infty}\gamma^*(z) = \gamma^{**}$ where $J'(\gamma^{**}) = -u^*$. Since $J(\cdot)$ is concave, this uniquely pins down the investment of the largest firms as a function of the asymptotical marginal value u^*. Equation (50) and $\lim_{z\to\infty} u_2'(z) = 0$ and then imply:

$$\left[r+\delta+g-\left(\rho\gamma^{**}+\frac{\rho^2}{2}\sigma^2\right)\right]u^* - J(\gamma^{**}) = 0.$$

Together, these conditions give

$$u^* \quad = \quad = \frac{J(\gamma^{**})}{r+\delta+g-\left(\rho\gamma^{**}+\frac{\rho^2}{2}\sigma^2\right)} = -J'(\gamma^{**}),$$

where the second equality follows from the first order condition corresponding to the firm's choice of process innovation rate. We can simplify this relation to

$$-\frac{J'(\gamma^{**})}{J(\gamma^{**})} = \frac{1}{\varpi-\rho\gamma^{**}}, \tag{52}$$

where I have defined $\varpi \equiv r+\delta+g-\frac{1}{2}\rho^2\sigma^2$.

To establish conditions that ensure this equation has a solution, note that the function $\mathsf{A}(\gamma) = J(\gamma) + (\varpi-\rho\gamma)J'(\gamma)$ is monotonically decreasing. Since we have $\mathsf{A}(0) = J(0) = 1$, the condition for there to be a unique equilibrium is that $\mathsf{A}(\overline{\gamma}) < 0$, which yields condition

$$\varpi - \rho\hat{\gamma} > -\frac{J(\overline{\gamma})}{J'(\overline{\gamma})},$$

which gives us condition (18). Therefore, a unique γ^{**} exists.

Now, in additition to condition $u_2(z_o) = 0$, we have the boundary conditions for large firms in Equation (51). Together, these three boundary conditions would allow us to fully characterize the solution to the 2nd order differential equation with a free boundary value z_o. Note that since we established the existence of γ^{**}, we have shown that the differential equation has a fixed point.

Next, we should check the stability of the fixed point. Assume that a solution exists and that around the equilibrium point $(u_1, u_2)^t = (u^*, u^*)^t$, it behaves as follows

$$u_1(z) = u^* + Be^{-\beta z}, \qquad u_2(z) = u^* - (\beta-1)Be^{-\beta z}, \qquad \text{for large } z, \tag{53}$$

which automatically satisfies (49). Letting $\gamma^*(z) = \gamma^{**} + \Delta\gamma(z)$ where the second term is small for large z, from Equation (48) we find $\Delta\gamma(z) = \frac{\beta-1}{J''(\gamma^{**})}Be^{-\beta z}$. The behavior of Equation (50) around the equilibrium point can be expressed as

$$u_2'(z) \approx \frac{2}{\rho^2\sigma^2}\left[r+\delta-\left(\rho\gamma^{**}-g+\frac{1}{2}\rho^2\sigma^2\right)(\beta-1)\right]Be^{-\beta z}. \tag{54}$$

We find the following quadratic equation that characterizes β

$$\frac{1}{2}\rho^2\sigma^2\beta(\beta-1) + \left(\rho\gamma^{**}-g+\frac{1}{2}\sigma^2\right)(\beta-1) - (r+\delta) = 0.$$

The condition for uniqueness is that this equation should have one positive and one strictly negative root, which is ensured by $r+g+\delta > \rho\gamma^{**} + \frac{1}{2}\rho^2\sigma^2$.[52] Since $\gamma^{**} < \overline{\gamma}$, this condition is subsumed by Equation (18). This establishes the existence of a stable, fixed point for the system of differential equations.

To derive the approximation in Section **??**, simply extend the form of the solution given by Equations 53. The smooth pasting condition requires $u_2(z_o) = 0$ and we find $u^* = (\beta - 1) Be^{-\beta z_o}$, while the value matching condition gives us $u_1(z_o) = \frac{1}{r} = u^*\left(1 + \frac{1}{\beta-1}\right) e^{z_o}$. □

Proposition. *See Proposition 2.*

Proof. I will peform the derivations under the assumption that γ_e is large enough so that all entry attempts are successful. The derivation in the other case is also very similar.

As with the strategy used in the case of the HJB equation, I begin by mapping the 2nd order differential equation into a first-order one defined in the two-dimensional space of $\boldsymbol{h}(z) = (h_1(z), h_2(z))' \equiv (H(z), h(z))'$. We can rewrite the stationary KFE (20) as

$$h_1'(z) = \boldsymbol{\mathcal{P}}_1(\boldsymbol{h}) = h_2, \tag{55}$$

$$h_2'(z) = \boldsymbol{\mathcal{P}}_2(\boldsymbol{h}, z) = \frac{2}{\rho^2\sigma^2}\left[\lambda_o - \lambda_e F_a\left(\frac{h_1(z) - \eta\tilde{z} - \rho\gamma_e}{1-\eta}\right) + (g_M + \delta) h_1(z) + (\rho\gamma^*(z) - g) h_2(z)\right] \tag{56}$$

If a stationary distribution exists, the system of differential equations above along with boundary conditions $h_1(z_o) = h_2(z_o) = 0$ and $\lim_{z\to\infty} h_1(z) = 1$ together pin down the solution and the unknown parameter λ_o.

Once again, to ensure that such a solution exists and is unique, we first examine the behavior of the system of differential equations around its equilibrium point $\boldsymbol{h} = (1,0)^t$ where $\frac{d}{dz}\boldsymbol{h}(z) = 0$. I linearize the system around this point to find

$$\Delta\boldsymbol{\mathcal{P}} \equiv \begin{bmatrix} 0 & 1 \\ \mathcal{P}_{21} & \mathcal{P}_{22} \end{bmatrix},$$

where, assuming $\eta > 0$ or $f_a(1) = 0$, we have

$$\mathcal{P}_{21} = \frac{2(g_M + \delta)}{\rho^2\sigma^2}, \qquad \mathcal{P}_{22} = \frac{2(\rho\gamma^* - g)}{\rho^2\sigma^2}.$$

For the stationary distribution to exist, this matrix needs to have a positive and a negative eigenvalue which is smaller than -1. The eigvenvalues are

$$\frac{\rho\gamma^* - g}{\rho^2\sigma^2} \pm \sqrt{\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 + \frac{2(g_M + \delta)}{\rho^2\sigma^2}}.$$

The product of the two is negative so long as $g_M + \delta > 0$. For the eigenvalue to be less than -1, we have the condition

$$g_M + \delta > \rho\gamma^* - g + \frac{1}{2}\rho^2\sigma^2.$$

To see this, note that the Pareto tail index is given by

$$\zeta \equiv \sqrt{\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 + \frac{2(g_M + \delta)}{\rho^2\sigma^2}} - \frac{\rho\gamma^* - g}{\rho^2\sigma^2} > 1, \tag{57}$$

suggesting

$$\left(\frac{\rho\gamma^* - g}{\rho^2\sigma^2}\right)^2 + \frac{2(g_M + \delta)}{\rho^2\sigma^2} > \left(\frac{\gamma^* - g}{\sigma^2} + 1\right)^2,$$

[52] Another way to see this is to note that we can simply linearize the system around the infinite limit of z. The linearized equation near $(u_1, u_2)^t = (u^*, u^*)^t$ is given by $\boldsymbol{u}'(z) = \Delta\boldsymbol{Q}\,\boldsymbol{u}(z)$ with $\Delta\boldsymbol{Q} \equiv \begin{bmatrix} -1 & 1 \\ Q_{21} & Q_{22} \end{bmatrix}$, where $Q_{21} = \frac{2(r+\delta)}{\sigma^2}$, $Q_{22} = -\frac{2(\rho\gamma^* - g)}{\rho^2\sigma^2} - 1$. If the determinant $-(Q_{22} + Q_{11}) < 0$ is negative, there will be two eigenvalues one positive and one negative. The condition for the determinant to be negative is $r + g + \delta > \rho\gamma^{**} + \frac{1}{2}\rho^2\sigma^2$.

and

$$\begin{aligned} \frac{2\left(g_M+\delta\right)}{\rho^2\sigma^2} &> \left(\frac{\rho\gamma^*-g}{\rho^2\sigma^2}+1\right)^2-\left(\frac{\rho\gamma^*-g}{\rho^2\sigma^2}\right)^2 \\ &= 2\left(\frac{\rho\gamma^*-g}{\rho^2\sigma^2}+\frac{1}{2}\right). \end{aligned}$$

□

Proposition. *See Proposition 3.*

Proof. I use Lemma 2 below that decomposes average productivity growth everywhere along the path. Consider a stationary balanced growth path, along which $F_t(\theta) = H(z_t(\theta))$. We also have $J_t(\theta) = J(\gamma^*(z_t(\theta)))$, which implies

$$\begin{aligned} \varepsilon^{J,1}_{\theta,t} &= \frac{\theta J_t'}{J_t} = \rho\frac{J'}{J} = \rho\varepsilon^J_1, \\ \varepsilon^{J,2}_{\theta,t} &= \frac{\theta^2 J_t''}{J_t} = \frac{\theta^2}{J_t}\left(\frac{\rho}{\theta}J'\right)' = \rho^2\frac{J''}{J} - \rho\frac{J'}{J} = \rho^2\left[\left(\varepsilon^J_1\right)^2+\varepsilon^J_2\right]-\rho\varepsilon^J_1, \\ \frac{\dot J_t}{J_t} &= -\rho g_\theta\frac{J'}{J} = -\rho g_\theta\varepsilon^J_1, \end{aligned}$$

where, with some abuse of notation, I have defined $J(z) \equiv J(\gamma^*(z))$. We then find

$$\begin{aligned} \rho g_\theta &= \lambda_o\left(\frac{\mathbb{E}_F\left[e^z J\right]}{\mathbb{E}_H\left[e^z J\right]} - \frac{e^{z^o}}{\mathbb{E}_H\left[e^z J\right]}\right) - g_M\left(1-\frac{\mathbb{E}_F\left[e^z J\right]}{\mathbb{E}_H\left[e^z J\right]}\right) + \mathbb{E}_\nu\left[\gamma^*\left(1+\varepsilon^J_1\right)\right] \\ &\quad + \frac{\sigma^2}{2}\left(\rho(\rho-1)+\rho^2\mathbb{E}_\nu\left[2\varepsilon^J_1+\left(\varepsilon^J_1\right)^2+\varepsilon^J_2\right]-\rho\mathbb{E}\left[\varepsilon^J_1\right]\right). \end{aligned}$$

□

Lemma 2. *(Evolution of Sectoral State Variables) Assume that (1) Gibrat's law holds for the largest firms, $\lim_{\theta\to\infty}\gamma_t(\theta) = \gamma^{**} < \infty$, and (2) the distribution of firm size has a finite expectation at all times, i.e., $m_t(\theta) = O(\theta^{-1-\rho})$ for all t. Consider optimal firm incumbent and entrant policies, that is, the paths of investment policies and exit threshold $(\gamma^*_t(\cdot), \theta_{o,t})$ and the rates of entry $\lambda_{e,t}$. Assume that the distribution of entrants is given by $f_{e,t}(\cdot)$. Let $f_t(\cdot)$ denote the distribution of firm productivity θ at time t, and let $\mathbb{E}_{\nu,t}$ denote expectation with respect to the share of each type in sales at time t.*[53] *The evolution of the sector state variables satisfies the following equations*

$$\frac{\dot{\overline{M}}_t}{\overline{M}_t} = \lambda_{e,t}-\lambda_{o,t}, \tag{58}$$

$$\begin{aligned} \frac{\dot{\overline{\theta}}_t}{\overline{\theta}_t} &= \frac{\lambda_{o,t}}{\rho}\left[\left(\frac{\overline{\theta}_{e,t}}{\overline{\theta}_t}\right)^\rho - \left(\frac{\theta_{o,t}}{\overline{\theta}_t}\right)^\rho\right] - \frac{\lambda_{e,t}-\lambda_{o,t}}{\rho}\left[1-\left(\frac{\overline{\theta}_{e,t}}{\overline{\theta}_t}\right)^\rho\right] + \mathbb{E}_{\nu,t}\left[\gamma_{\theta,t}\left(1+\frac{1}{\rho}\epsilon^{J,1}_{\theta,t}\right)\right] \\ &\quad + (\rho-1)\frac{\sigma^2}{2}\left(1+\frac{1}{\rho-1}\mathbb{E}_{\nu,t}\left[2\varepsilon^{j,1}_{\theta,t}+\frac{\varepsilon^{J,2}_{\theta,t}}{\rho}\right]\right) + \frac{1}{\rho}\mathbb{E}_{\nu,t}\left[\frac{\dot J_t}{J_t}\right], \end{aligned} \tag{59}$$

where $\lambda_{o,t}$ as in Equation (43), I have defined and $J_t(\theta) \equiv J(\gamma_{\theta,t}(\theta))$ and the corresponding elasticities $\varepsilon^{J,1}_{\theta,t} \equiv \frac{\theta\partial J_t}{J_t\partial\theta}$ and $\varepsilon^{J,2}_{\theta,t} \equiv \frac{\theta^2\partial^2 J_t}{J_t\partial\theta^2}$.

Proof. Let us calculate the integral in the definition of the aggregate productivity in Equation (47) using the KFE (45):

$$\frac{d}{dt}\left(\overline{M}_t\overline{\theta}^\rho_t\right) \overset{(1)}{=} \overline{M}_t\lambda_{e,t}\int d\theta\, f_{e,t}\,\theta^\rho J_t - \int d\theta\,\frac{\partial}{\partial\theta}\left[\gamma_t\,\theta m_t\right]\theta^\rho J_t + \frac{\sigma^2}{2}\int d\theta\,\frac{\partial^2}{\partial^2\theta}\left[\theta^2 m_t\right]\theta^\rho J_t + \int d\theta\, m_t\theta^\rho\dot J_t,$$

[53] Note that that the notation in this proof is slightly different from the one in the main text, to allow for recylcing variable names.

$$
\begin{aligned}
&\overset{(2)}{=} \overline{M}_t \lambda_{e,t}\, \mathbb{E}_{f_{e,t}}\left[\theta^\rho J_t\right] - \left[\gamma_t\, m_t \theta^{1+\rho} J_t\right]_{\theta_{o,t}}^{\infty} + \rho \int d\theta\, \gamma_t\, m_t \theta^\rho J_t \left(1 + \frac{\theta J_t'}{\rho J_t}\right) \\
&\quad + \frac{\sigma^2}{2}\left[\frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\theta^\rho J_t\right]_{\theta_t^o}^{\infty} - \frac{\rho\sigma^2}{2}\int d\theta\, \frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\left(\theta^{\rho-1} J_t + \frac{1}{\rho}\theta^\rho J_t'\right) + \overline{M}_t\, \mathbb{E}_{f_t}\left[\theta^\rho \dot{J}_t\right], \\
&\overset{(3)}{=} \overline{M}_t \lambda_{e,t}\, \mathbb{E}_{f_{e,t}}\left[\theta^\rho J_t\right] + \overline{M}_t\, \mathbb{E}_{f_t}\left[\gamma_{\theta,t}\theta^\rho J_t\left(\rho + \varepsilon_{\theta,t}^{J,1}\right)\right] - \frac{\sigma^2}{2}\theta_{o,t}^\rho \frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\Big|_{\theta_{o,t}} + \overline{M}_t\, \mathbb{E}_{f_t}\left[\theta^\rho \dot{J}_t\right] \\
&\quad - \frac{\rho\sigma^2}{2}\left[\theta^2 m_t\left(\theta^{\rho-1} J_t + \frac{1}{\rho}\theta^\rho J_t'\right)\right]_{\theta_{o,t}}^{\infty} \\
&\quad + \rho\left(\rho-1\right)\frac{\sigma^2}{2}\int d\theta\, m_t \theta^\rho J_t\left(1 + \frac{2}{\rho-1}\frac{\theta J_t'}{J_t} + \frac{1}{\rho\left(\rho-1\right)}\frac{\theta^2 J_t''}{J_t}\right) \\
&\overset{(4)}{=} \overline{M}_t \lambda_{e,t}\, \mathbb{E}_{f_{e,t}}\left[\theta^\rho J_t\right] + \overline{M}_t\, \mathbb{E}_{f_t}\left[\gamma_{\theta,t}\theta^\rho J_t\left(\rho + \varepsilon_{\theta,t}^{J,1}\right)\right] - \frac{\sigma^2}{2}\theta_{o,t}^\rho \frac{\partial}{\partial\theta}\left(\theta^2 m_t\right)\Big|_{\theta_{o,t}} + \overline{M}_t\, \mathbb{E}_{f_t}\left[\theta^\rho \dot{J}_t\right] \\
&\quad + \frac{\sigma^2}{2}\, \mathbb{E}_{f_t}\left[\theta^\rho J_t\left(\rho\left(\rho-1\right) + 2\rho\varepsilon_{\theta,t}^{J,1} + \varepsilon_{\theta,t}^{J,2}\right)\right]
\end{aligned}
$$

where in equality (1), I have substituted for $\frac{\partial m_t}{\partial t}$ from Equation (46), in equality (2), used the chain rule to expand the integrals, and in equality (3), used the fact that $m_t\left(\theta\right) = O\left(\theta^{-(1+\rho)}\right)$ to drop the boundary terms involving $\theta \to \infty$. Finally, in equality (4) I have again used $m_t\left(\theta\right) = O\left(\theta^{-(1+\rho)}\right)$ and $m_t\left(\theta_{o,t}\right) = 0$ to drop the term involving boundary values.

Dividing the two sides of the last equation by $\overline{M}_t \overline{\theta}_t^\rho$, we find

$$
\begin{aligned}
\frac{\dot{\overline{M}}_t}{\overline{M}_t} + \rho\frac{\dot{\overline{\theta}}_t}{\overline{\theta}_t} &= \lambda_{e,t}\left(\frac{\overline{\theta}_{e,t}}{\overline{\theta}_t}\right)^\rho - \lambda_{o,t}\left(\frac{\theta_{o,t}}{\overline{\theta}_t}\right)^\rho + \rho\mathbb{E}_{s,t}\left[\gamma_{\theta,t}\left(1 + \epsilon_{\theta,t}^{J,1}\right)\right] \\
&\quad + \frac{\sigma^2}{2}\left(\rho\left(\rho-1\right) + \mathbb{E}_{s_t}\left[2\rho\varepsilon_{\theta,t}^{J,1} + \varepsilon_{\theta,t}^{J,2}\right]\right) + \mathbb{E}_{s_t}\left[\frac{\dot{J}_t}{J_t}\right].
\end{aligned}
$$

Using $\dot{\overline{M}}_t / \overline{M}_t = \lambda_{e,t} - \lambda_{o,t}$ gives us (59). □

7.2 Section 3

Proposition. *The result generalizes Proposition 4 of the main text, by assuming a generalized industry-wide stock of knowledge*

$$
\tilde{\theta}_t^\vartheta = \frac{1}{\overline{M}_t}\int \theta^\vartheta m_t\left(\theta\right) d\theta, \qquad \vartheta \neq 0,
$$

and additional shocks to the productivity of entrants, with a log-Normal distribution.

Proof. The proof proceeds in the following steps. I will first discretize the probelm. I then set up the social planner's problem in the discretized space, and derive the first order conditions. Finally, I combine the first order conditions and take the continuous time limit to characterize the solution.

Discretization I discretize the space of time and states following the random walk construction of Brownian motion processes (Dixit and Pindyck, 1994). Using Itô's lemma, we find

$$
d\log\theta_t = \left(\gamma_t - \frac{1}{2}\sigma^2\right)dt + \sigma d\mathcal{W}_t.
$$

Abusing some notation to recycle variable names, for the sake of this proof I will define $t \equiv j\Delta t$, $\rho\ln\theta_t \equiv z\Delta h$, where $\Delta h = \rho\sigma\sqrt{\Delta t}$, to map the stochastic process for the evolution of profitability states θ_t into a random walk over a space $z \in \mathbb{Z}$ and $j \in \mathbb{N}\cup\{0\}$. The probability of moving from z by ± 1 between j and $j+1$ is given by

$$
\mu_j^+\left(z\right) = \frac{1}{2}\left[1 + \left(\gamma_j\left(z\right) - \frac{1}{2}\sigma^2\right)\frac{\sqrt{\Delta t}}{\sigma}\right],
$$

$$\mu_j^-(z) = \frac{1}{2}\left[1 - \left(\gamma_j(z) - \frac{1}{2}\sigma^2\right)\frac{\sqrt{\Delta t}}{\sigma}\right],$$

where I have let $\gamma_j(z) \equiv \gamma_t(\theta_t)$.

Problem Define $M_j(z) \equiv \Delta h \sum_{z'=z_{o,j}}^{z} m_j(z')$ and let $\tilde{z}_j$ be defined through

$$e^{\tilde{z}_j} \equiv \left(\Delta h \sum_{z=z_{o,j}}^{\infty} \frac{m_j(z)}{\overline{M}_j} e^{\frac{\vartheta}{\rho} z\Delta h}\right)^{\frac{\rho}{\vartheta}}, \tag{60}$$

corresponding to the industry-wide index of knowledge. The social planner's problem can be written as that of maximizing

$$\Delta t\, N_j \sum_{j=0}^{\infty} e^{-jr\Delta t} \log q_j,$$

subject to the constraints

$$N_j q_j = \left(\Delta h \sum_{z=z_{o,j}}^{\infty} m_j(z)\, e^{\frac{z\Delta h}{1+\rho}}\, l_j(z)^{\frac{\rho}{1+\rho}}\right)^{\frac{1+\rho}{\rho}}, \tag{61}$$

$$N_j = L_{e,j} + \Delta h \sum_{z=z_{o,j}}^{\infty} m_j(z)\left[\psi_f + l_j(z) + \bar{i}_j\right],$$

$$\overline{M}_j \bar{i}_j = \Delta h \sum_{z=z_{o,j}}^{\infty} m_j(z)\, l_j(z)\, \varphi(\gamma_j(z)), \tag{62}$$

$$m_{j+1}(z) = \frac{L_{e,j}}{\psi_e}\Delta t\, f_j(z) + m_j(z+1)\,\mu_j^-(z+1) + m_j(z-1)\,\mu_j^+(z-1),$$

$$f_j(z) = \frac{\Delta h}{\rho}\sum_{z_a=z_{o,j}}^{\infty} f_e\left(\frac{z - \eta\tilde{z}_j - (1-\eta)z_a}{\rho}\Delta h\right)\frac{m_j(z_a)}{\overline{M}_j} f_a\left(\frac{M_j(z_a)}{\overline{M}_j}\right), \tag{63}$$

where $\bar{i}_j$ denotes the average innovation workers of the incumbent firms. Let the Lagrange multipliers corresponding to the three constraints above be denoted by $\Delta t\,\chi_j e^{-jr\Delta t}$, $\Delta t\,\omega_j e^{-jr\Delta t}$, $\Delta t\,\omega_j^i e^{-jr\Delta t}$, $\Delta h\,\nu_{j+1}(z)\,e^{-jr\Delta t}$, respectively. The social planner decides on the per-capita consumption q_j, on the exit threshold at time $j\Delta t$, which I denote by $z_{o,j}$, on the total number of innovation workers in the entrepreneurial work $L_{e,j}$ and the average number of innovation workers in incumbent firms $\bar{i}_j$ at each point in time, the number of production workers assigned to each firm $l_j(z)$, the innovation investments of firms $\gamma_j(z)$, and the mass of firms in each state $m_j(z)$.

First Order Conditions (I): consumption, incumbent innovation and production The FOC with respect to q_j gives $\frac{N_j}{q_j} = \chi_j N_j$, which yields $q_j\chi_j = 1$. The FOC with respect to the average number of incumbent innovation workers $\bar{i}_j$ is given by $\omega_j^i = \omega_j$. First, the FOC with respect to $l_j(z)$ gives

$$\left(\frac{N_j q_j e^{z\Delta h}}{l_j(z)}\right)^{\frac{1}{1+\rho}} \chi_j = \omega_j\left(1 + \varphi(\gamma_j(z))\right). \tag{64}$$

Substituting $\chi_j = 1/q_j$, using the definition of function $J(\gamma) \equiv (1+\varphi(\gamma))^{-\rho}$, and letting Θ_j be defined through

$$\overline{M}_j\Theta_j^\rho = \Delta h \sum_{z=z_j^o}^{\infty} m_j(z)\, e^{z\,\Delta h}\, J(\gamma_j(z)),$$

we can substitute for $l_j(z)$ in Equation (61) to find $\chi_j\Theta_j\overline{M}_j^{\frac{1}{\rho}} = \omega_j$.

First Order Condition (II): The FOC with respect to $\gamma_j(z)$ we find

$$\frac{1}{2}\frac{\sqrt{\Delta t}}{\sigma}\left[v_{j+1}(z+1)-v_{j+1}(z-1)\right] = \Delta t\, l_j(z)\;\omega_j\,\varphi'\left(\gamma_j(z)\right), \tag{65}$$

$$= \Delta t\,\chi_j\left(\frac{\chi_j}{\omega_j}\right)^{\rho} N_j q_j\, e^{z\Delta h}\left(1+\varphi\left(\gamma_j(z)\right)\right)^{-(1+\rho)}\;\varphi'\left(\gamma_j(z)\right),$$

$$= -\Delta t\,\frac{N_j}{\overline{M}_j}\frac{e^{z\Delta h}}{\Theta_j}J'\left(\gamma_j(z)\right), \tag{66}$$

where, in the second line, I have again substitute for $l_j(z)$ from Equation (61).

First Order Conditions (III): social free entry The FOC with respect to $L_{e,j}$ gives

$$\omega_j\psi_e = \Delta h\sum_{z=z_{o,j}}^{\infty} f_j(z)\, v_{j+1}(z). \tag{67}$$

First Order Conditions (IV): incumbent value The FOC with respect to $m_{j+1}(z)$ is the most involved calculation. The first term in Equation (61) gives the flow social value of production for a firm z at time interval $j+1$:

$$\Delta t\Delta h\,\chi_j\,\frac{1+\rho}{\rho}\left(N_j q_j\right)^{\frac{1}{1+\rho}} e^{\frac{z\Delta h}{1+\rho}}\, l_j(z)^{\frac{\rho}{1+\rho}} = \Delta t\Delta h\,\chi_j\,\frac{1+\rho}{\rho}\,N_j q_j\left(\frac{\chi_j}{\omega_j}\right)^{\rho} e^{z\Delta h}\left(1+\varphi\left(\gamma_j(z)\right)\right)^{-\rho},$$

$$= \Delta t\Delta h\,\frac{1+\rho}{\rho}\,\frac{N_j}{\overline{M}_j}\frac{e^{z\Delta h}}{\Theta_j}J\left(\gamma_j(z)\right).$$

The second and third terms give the employment costs by

$$\Delta t\Delta h\left[\omega_j\,\psi_f+\omega_j\,l_j(z)\,\left(1+\varphi\left(\gamma_j(z)\right)\right)\right] = \Delta t\Delta h\left[\omega_j\,\psi_f+l_j(z)\left(\frac{N_j q_j e^{z\Delta h}}{l_j(z)}\right)^{\frac{1}{1+\rho}}\chi_j\right],$$

$$= \Delta t\Delta h\left[\omega_j\,\psi_f+\frac{N_j}{\overline{M}_j}\frac{e^{z\Delta h}}{\Theta_j}J\left(\gamma_j(z)\right)\right],$$

where on the second line, I have substituted from Equation (64). The fourth term gives

$$\Delta h\left[e^{r\Delta t}v_j(z)-\frac{L_{e,j}}{\psi_e}\left(\Delta h\sum_{x=z_{o,j}}^{\infty}\frac{\partial f_j(x)}{\partial m_j(z)}v_{j+1}(x)\right)\Delta t+\mu^{+}_{j+1}(z)\,v_{j+1}(z+1)+\mu^{-}_{j+1}(z)\,v_{j+1}(z-1)\right].$$

The key term is the second term in the expression above, that captures the knowledge spillovers from incumbents to entrants. This spillover term is in turn composed of two components: first, the contribution of incumbent firms to the average stock of knowledge $\tilde{z}$, and second, the direct diffusion of ideas of incumbents to entrants through adoption $f_a(\cdot)$. Defining $H_j(z)\equiv\Delta h\sum_{z'\leq z}m_j(z)/\overline{M}_j$ and $h_j(z)\equiv m_j(z)/\overline{M}_j$ to simplify notation, we can computing the first term as

$$-\frac{\eta}{\rho}\frac{\partial\tilde{z}_j}{\partial m_j(z)}\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty}v_{j+1}(x)\;f'_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)\,z_a}{\rho}\Delta h\right)h_j(z_a)\;f_a\left(H_j(z_a)\right).$$

Taking the derivative of expression (60), we find

$$\frac{\partial e^{\tilde{z}_j}}{\partial m_j(z)} = \frac{\Delta h}{\overline{M}_j}\frac{\rho}{\vartheta}\left[e^{\frac{\vartheta}{\rho}\left(z\Delta h-\tilde{z}_j\right)}-1\right]\times e^{\tilde{z}_j}.$$

Similarly, the second term may be computed as

$$\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty} v_{j+1}(x)\, f_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right)\left[\frac{\partial h_j(z_a)}{\partial m_j(z)} f_a\left(H_j(z_a)\right)+h_j(z_a)\frac{\partial H_j(z_a)}{\partial m_j(z)} f_a'\left(H_j(z_a)\right)\right].$$

Calculating the derivatives within the square brackets yields

$$\begin{aligned}&\frac{1}{\overline{M}_j}\left[\left(\mathbb{I}(z_a=z_j)-\frac{m_j(z_a)}{\overline{M}_j}\Delta h\right) f_a\left(H_j(z_a)\right)+h_j(z_a)\left(\mathbb{I}(z_a\geq z_j)-\frac{M_j(z_a)}{\overline{M}_j}\right)\Delta h\, f_a'\left(H_j(z_a)\right)\right]\\ &\qquad=\frac{1}{\overline{M}_j}\left[\left(\mathbb{I}(z_a=z)-h_j(z_a)\,\Delta h\right) f_a\left(H_j(z_a)\right)+h_j(z_a)\left(\mathbb{I}(z_a\geq z)-H_j(z_a)\right)\Delta h\, f_a'\left(H_j(z_a)\right)\right]\quad.\end{aligned}$$

Combining all these expressions, we can write the FOC with respect to $m_{j+1}(z)$ as

$$\begin{aligned}e^{r\Delta t}v_j(z) &= \left[\frac{1}{\rho}\frac{N_{j+1}}{\overline{M}_{j+1}}\frac{e^{z\Delta h}}{\Theta_{j+1}^{\rho}} J\left(\gamma_j(z)\right)-\psi_f\,\omega_{j+1}+\lambda_{e,j}\left(\tilde{u}_{s,j}(z)+u_{s,j}(z)\right)\right]\Delta t\\ &\qquad +\mu_{j+1}^{+}(z)\,v_{j+1}(z+1)+\mu_{j+1}^{-}(z)\,v_{j+1}(z-1),\end{aligned}\tag{68}$$

where $\lambda_{e,j}=L_{e,j}/\overline{M}_j\psi_e$ is the rate of entry, and the two spillover functions are given by

$$\tilde{u}_{s,j}(z) \equiv -\frac{\eta}{\vartheta}\left[e^{\frac{\vartheta}{\rho}(z\Delta h-\tilde{z}_j)}-1\right]\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty} v_{j+1}(x)\, f_e'\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right) h_j(z_a)\, f_a\left(H_j(z_a)\right),\tag{69}$$

$$\begin{aligned}u_{s,j}(z) \equiv\ & \frac{\Delta h}{\rho}\sum_{x=z_{o,j}}^{\infty} f_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z}{\rho}\Delta h\right) f_a\left(H_j(z)\right) v_{j+1}(x)\\ &-\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty} f_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right) h_j(z_a)\, f_a\left(H_j(z_a)\right) v_{j+1}(x)\\ &+\frac{\Delta h^2}{\rho}\sum_{z_a=z}^{\infty}\sum_{x=z_{o,j}}^{\infty} f_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right) h_j(z_a)\, f_a'\left(H_j(z_a)\right) v_{j+1}(x)\\ &-\frac{\Delta h^2}{\rho}\sum_{x=z_{o,j}}^{\infty}\sum_{z_a=z_{o,j}}^{\infty} f_e\left(\frac{x-\eta\tilde{z}_j-(1-\eta)z_a}{\rho}\Delta h\right) H_j(z_a)\, f_a'\left(H_j(z_a)\right) v_{j+1}(x).\end{aligned}\tag{70}$$

Limit of Continuous Time Now, let us take the limit of $\Delta t\to\infty$, define $\hat{\theta}\equiv\rho\log\theta$ and $\widehat{\tilde{\theta}}_t\equiv\rho\log\tilde{\theta}_t$, and with some further abuse of notation, let $v_{s,t}(\hat{\theta})$ denote the social value of the a firm with productivity $e^{\hat{\theta}/\rho}$. I will first compute the limiting expressions for the summations in Equations (69) and (70). For l define the auxiliary function

$$\mathcal{V}_t(\hat{\theta})\equiv\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}}{\rho}\right)\frac{dx}{\rho}.$$

Assuming $v_{s,t}(\hat{\theta}_{o,t})=0$ and the right tail of $f_e(x)$ falls faster than e^{-x}, using integration by parts, we find

$$\begin{aligned}\mathcal{V}_t'(\hat{\theta}) &= -\frac{1-\eta}{\rho}\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}(x)\, f_e'\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}}{\rho}\right)\frac{dx}{\rho},\\ &= -\frac{1-\eta}{\rho}\left[v_{s,t}(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}}{\rho}\right)\right]_{x=\hat{\theta}_{o,t}}^{\infty}\\ &\qquad +(1-\eta)\int_{\hat{\theta}_{o,t}}^{\infty} v_{s,t}'(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right)\frac{dx}{\rho},\end{aligned}$$

$$= (1-\eta)\int_{\hat{\theta}_{o,t}}^{\infty} \nu'_{s,t}(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right)\frac{dx}{\rho}.$$

Using a similar argument, we can write the limit of the first spillover term may be found

$$\begin{aligned}
\tilde{u}_{s,t}(\hat{\theta}) &= -\frac{\eta}{\vartheta}e^{\frac{\vartheta}{\rho}\left(\hat{\theta}-\widehat{\tilde{\theta}}_t\right)}\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty}\nu_{s,t}(x)\, f'_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right)h_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\frac{dx}{\rho}d\hat{\theta}_a,\\
&= \eta\frac{\rho}{\vartheta}e^{\frac{\vartheta}{\rho}\left(\hat{\theta}-\widehat{\tilde{\theta}}_t\right)}\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty}\nu'_{s,t}(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right)h_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\frac{dx}{\rho}d\hat{\theta}_a.
\end{aligned}$$

For the second term, the limit is given by

$$\begin{aligned}
u_{s,t}(\hat{\theta}) &= f_a(H_t(\hat{\theta}))\,\mathcal{V}_t(\hat{\theta})-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t(\hat{\theta}_a)\, h_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a\\
&\quad+\int_{\hat{\theta}}^{\infty}\mathcal{V}_t(\hat{\theta}_a)\, h_t(\hat{\theta}_a)\, f'_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a\\
&\quad-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t(\hat{\theta}_a)\, H_t(\hat{\theta}_a)\, f'_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a.
\end{aligned}$$

Once again, using integration by part, we can simplify this expression to find

$$\begin{aligned}
u_{s,t}(\hat{\theta}) &= \mathcal{V}_t(\hat{\theta})\, f_a(H_t(\hat{\theta}))-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}_t(\hat{\theta}_a)\, h_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a\\
&\quad+\left[\mathcal{V}_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\right]_{\hat{\theta}_a=\hat{\theta}}^{\infty}-\int_{\hat{\theta}}^{\infty}\mathcal{V}'_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a\\
&\quad-\left[\mathcal{V}_t(\hat{\theta}_a)\, H_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\right]_{x=\hat{\theta}_{o,t}}^{\infty}+\int_{\hat{\theta}_{o,t}}^{\infty}\frac{d}{d\hat{\theta}_a}\left[\mathcal{V}_t(\hat{\theta}_a)\, H_t(\hat{\theta}_a)\right] f'_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a,\\
&= -\int_{\hat{\theta}}^{\infty}\mathcal{V}'_t(\hat{\theta}_a)\, h_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a+\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}'_t(\hat{\theta}_a)\, H_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a,\\
&= \int_{\hat{\theta}_{o,t}}^{\hat{\theta}}\mathcal{V}'_t(\hat{\theta}_a)\, f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a-\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}'_t(\hat{\theta}_a)\left[1-H_t(\hat{\theta}_a)\right] f_a(H_t(\hat{\theta}_a))\,d\hat{\theta}_a,
\end{aligned}$$

where in the second equality, I have cancelled out the same terms that appear on the first three lines, and in the last equality, I have added and subtracted a term $\int_{\hat{\theta}_{o,t}}^{\infty}\mathcal{V}'_t f_a d\hat{\theta}_a$. We can now

$$\begin{aligned}
u_{s,t}(\hat{\theta}) &= (1-\eta)\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty}\nu'_{s,t}(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right) f_a(H_t(\hat{\theta}_a))\frac{dx}{\rho}d\hat{\theta}_a\\
&\quad-(1-\eta)\int_{\hat{\theta}_{o,t}}^{\infty}\int_{\hat{\theta}_{o,t}}^{\infty}\nu'_{s,t}(x)\, f_e\left(\frac{x-\eta\widehat{\tilde{\theta}}_t-(1-\eta)\hat{\theta}_a}{\rho}\right)\left[1-H_t(\hat{\theta}_a)\right] f_a(H_t(\hat{\theta}_a))\frac{dx}{\rho}d\hat{\theta}_a.
\end{aligned}$$

Putting everything together, we can combine the FOCs with respect to $\gamma_j(z)$ and $m_j(z)$ to find in the limit:

$$\begin{aligned}
r\, v_{s,t}(\hat{\theta})-\frac{\partial v_{s,t}(\hat{\theta})}{\partial t} &= \max_{\gamma}\ \frac{1}{\rho}\frac{N_t}{\overline{M}_t}\frac{e^{\hat{\theta}}}{\overline{\theta}_t^{\rho}}J(\gamma)-\psi_f\,\omega_t+\lambda_{e,t}\left(\tilde{u}_{s,t}(\hat{\theta})+u_{s,t}(\hat{\theta})\right)\\
&\quad+\left(\gamma-\frac{\rho}{2}\sigma_\theta^2\right)\frac{\partial v_{s,t}(\hat{\theta})}{\partial\hat{\theta}}+\frac{\rho^2\sigma_\theta^2}{2}\frac{\partial^2 v_{s,t}(\hat{\theta})}{\partial\hat{\theta}^2}.
\end{aligned}$$

Let us now assume that a social optimal equilibrium converges to a stationary constant growth path with growth rate $g_{\theta,s}$, where $\omega_t\to\omega$, $\overline{M}_t\to\overline{M}e^{g_N t}$, and $\overline{\theta}_t\to\overline{\theta}e^{g_\theta t}$. We can define $z_t(\hat{\theta})\equiv\hat{\theta}-\rho g_{\theta,s}t+\log\left(\frac{1}{\rho\psi_f\omega}\frac{N_0}{\overline{M}}\overline{\theta}^{-\rho}\right)$ and transform $v_{s,t}(\hat{\theta})=\omega\psi_f V(z_t(\hat{\theta}))$. Similar to the construction of the market equilibrium path, this implies that $z_t(\hat{\theta}_{o,t})$ and $z_t(\tilde{\theta}_t)$ also converge to constants $z_{o,s}$ and $\tilde{z}_s$ that are potentially different from the corresponding market equilibrium values.

$$r\,V_s\,(z) = \max_{\gamma} e^z J\,(\gamma) - 1 + \lambda_{e,s}\,(U_s\,(z) - \Psi_s) + \left(\gamma - \rho\left(g_{\theta,s} + \frac{1}{2}\sigma_\theta^2\right)\right) V_s'\,(z) + \frac{\rho^2}{2}\sigma_\theta^2\,V_s''\,(z)\,,$$

where

$$\begin{aligned}
U_s\,(z) &\equiv \eta\,\tilde{u}_s\,\frac{\rho}{\vartheta}\left[e^{\frac{\vartheta}{\rho}(z-\tilde{z})} - 1\right] + (1-\eta)\,(u_s\,(z) - \psi_s)\\
\tilde{u}_s &\equiv \int_{z_0}^{\infty} \mathcal{V}_e'\,(x)\;h_s\,(x)\;f_a\,(H_s\,(x))\,dx,\\
u_s\,(z) &\equiv \int_{z_0}^{z} \mathcal{V}_e'\,(x)\;f_a\,(H_s\,(x))\,dx,\\
\psi_s &\equiv \int_{z_0}^{\infty} \mathcal{V}_e'\,(x)\;[1 - H_s\,(x)]\;f_a\,(H_s\,(x))\,dx,\\
\mathcal{V}_e\,(z) &\equiv \int_{z_0}^{\infty} V\,(x)\;f_e\left(\frac{x - \eta\tilde{z} - (1-\eta)\,z}{\rho}\right)\frac{dx}{\rho}.
\end{aligned}$$

In the particular case where γ_e has an atomic distribution, we have $\mathcal{V}_e\,(z) \equiv V\,(\eta\tilde{z} + (1-\eta)\,z + \rho\gamma_e)$. □

7.3 Section 4

Specification of the Innovation Cost Function With the functional form assumed in Equation (39), the condition for the concavity of $J\,(\cdot)$ from Equation (5) gives

$$\kappa > (1+\rho)\,(1+\kappa)\,\frac{\varphi\,(\gamma)}{1+\varphi\,(\gamma)}.$$

It then follows that an upper bound for the rate of productivity growth given in Equation (40) ensures the convavity of function $J\,(\cdot)$. Throughout, I assume the value given by Equation (40) as the maximum attainable rate of productivity growth for firms.

In order to calibrate parameter κ, I rely on the estimates of the user cost elasticity of R&D that are available in the empirical literature. We can compute a counterpart for this elasticity based on the model by noting that $i = l\,\varphi\,(\gamma)$ and finding the elasticity of the innovation costs with respect to a proportional change $(1+\tau)$ in innovation costs. We have

$$\begin{aligned}
\frac{1+\tau}{i}\frac{\partial i}{\partial\,(1+\tau)} &= \frac{1+\tau}{l}\frac{\partial l}{\partial\,(1+\tau)} + \frac{1+\tau}{\varphi\,(\gamma)}\frac{\partial\varphi\,(\gamma)}{\partial\,(1+\tau)},\\
&= \left[1 - (1+\rho)\,\frac{\varphi}{1+\varphi}\right]\frac{1+\tau}{\varphi\,(\gamma)}\frac{\partial\varphi\,(\gamma)}{\partial\,(1+\tau)},
\end{aligned}$$

where I have used the fact that $l^{-\frac{1}{1+\rho}} \propto 1+\varphi$. Focusing on the limit of large firms, we can compute the elasticity of $\varphi\,(\gamma^{**})$ with respect to $1+\tau$ using the condition

$$\frac{\rho}{\gamma^{**}}\frac{\varphi}{1+\varphi}\,(1+\tau)\,(1+\kappa) = \frac{1}{r^{**}},$$

where r^{**} is the effective discount rate, which is not influenced by this partial equilibrium analysis. We then find

$$\left[1 - \frac{\varphi}{1+\varphi} - \frac{1}{1+\kappa}\right]\frac{1+\tau}{\varphi\,(\gamma)}\frac{\partial\varphi\,(\gamma)}{\partial\,(1+\tau)} = -1.$$

Combining the two expressions above, we have

$$\frac{1+\tau}{i}\frac{\partial i}{\partial\,(1+\tau)} = -\,(1+\kappa)\,\frac{1 - (1+\rho)\,\frac{\varphi}{1+\varphi}}{\kappa - \frac{\varphi}{1+\varphi}}.$$

Two-Sector Extension of the Model For the exercise performed in Section 4, I include data from two sectors manufacturing and retail trade, where the employment in each sector grows at a different rate. Below, I generalize to account for two sectors and show that the calibration strategy of Section 4 is compatible with the model.

Define the per-capita real consumption at time t as a function of per-capita consumption of manufacturing goods $q_{\mathrm{M},t}$ and retail goods $q_{\mathrm{R},t}$ as follows

$$q_t = \left(q_{\mathrm{M},t}^{\frac{\varsigma-1}{\varsigma}} + q_{\mathrm{R},t}^{\frac{\varsigma-1}{\varsigma}} \right)^{\frac{\varsigma}{\varsigma-1}},$$

where $\varsigma \in (0,1)$ is the consumer's elasticity of substitution between the products of the two sector.[54]

$$P_t = \left(P_{\mathrm{M},t}^{1-\varsigma} + P_{\mathrm{R},t}^{1-\varsigma} \right)^{\frac{1}{1-\varsigma}},$$

where Once again, normalizing the per-capita expenditure of households to unity, we find sectoral expenditures to be

$$X_{\mathrm{M},t} = P_{\mathrm{M},t} q_{\mathrm{M},t} = \left(\frac{P_{\mathrm{M},t}}{P_t} \right)^{1-\varsigma}, \qquad X_{\mathrm{R},t} = P_{\mathrm{R},t} q_{\mathrm{R},t} = \left(\frac{P_{\mathrm{R},t}}{P_t} \right)^{1-\varsigma}.$$

The relative price indices of the two sectors are given by

$$\frac{P_{\mathrm{M},t}}{P_{\mathrm{R},t}} = \frac{\bar{\theta}_{\mathrm{R},t} \overline{M}_{\mathrm{R},t}^{\frac{1}{\rho}}}{\bar{\theta}_{\mathrm{M},t} \overline{M}_{\mathrm{M},t}^{\frac{1}{\rho}}},$$

which gives us the following relationship between the per-capita expenditure between the two sectors

$$\frac{X_{\mathrm{M},t}}{X_{\mathrm{R},t}} = \left(\frac{\bar{\theta}_{\mathrm{R},t}}{\bar{\theta}_{\mathrm{M},t}} \right)^{1-\varsigma} \left(\frac{\overline{M}_{\mathrm{R},t}}{\overline{M}_{\mathrm{M},t}} \right)^{\frac{1-\varsigma}{\rho}}.$$

In the data, the number of firms grow faster in retail and the sectoral productivity grows faster in manufacturing. So long as the differences in the rates of growth in the number of firms and aggregate productivity between the two sectors satisfies

$$g_{\theta,\mathrm{M}} - g_{\theta,\mathrm{R}} > \frac{1}{\rho} \left(g_{M,\mathrm{R}} - g_{M,\mathrm{M}} \right),$$

we will find that the expenditure share of manufacturing asymptotically converges to zero at the rate

$$(1-\varsigma) \left[g_{\theta,\mathrm{M}} - g_{\theta,\mathrm{R}} - \frac{1}{\rho} \left(g_{M,\mathrm{R}} - g_{M,\mathrm{M}} \right) \right].$$

It then follows that we can construct a general equilibrium path where wages converge to a constant, the total expenditure in retail sector grows at the rate $g_{M,\mathrm{R}} = g_N$, the total expenditure in manufacturing grows at the rate given by

$$g_N - (1-\varsigma) \left[g_{\theta,\mathrm{M}} - g_{\theta,\mathrm{R}} - \frac{1}{\rho} \left(g_{M,\mathrm{R}} - g_{M,\mathrm{M}} \right) \right].$$

The mass of firms in each sector grows at the same rate as the rate of growth of the total expenditure of consumers in the sector.

This shows that the model is compatible with the exercise used in Section 4, where each sector is identified with a distinct rate of growth in the mass of firms (or alternatively, sectoral employment) and productivity. We can pin down the free parameter ς based on the observed rates of growth in the mass of firms and productivity.

[54] Based on the available empirical evidence, the empirically relevant case is one in which the expenditure shares of sectors with falling prices also fall, suggesting that sectoral goods are gross complements (Comin et al., 2015).