Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 15-020

"Growth through Heterogeneous Innovations Second Version"

by

Ufuk Akcigit, and William R. Kerr

Growth through Heterogeneous Innovations*

Ufuk Akcigit
University of Pennsylvania and NBER

William R. Kerr
Harvard University and NBER

March 25, 2015

Abstract

We study how external versus internal innovations promote economic growth through a tractable endogenous growth framework with multiple innovation sizes, multi-product firms, and entry/exit. Firms invest in external R&D to acquire new product lines and in internal R&D to improve their existing product lines. A baseline model derives the theoretical implications of weaker scaling for external R&D versus internal R&D, and the resulting predictions align with observed empirical regularities for innovative firms. Quantifying a generalized model for the recent U.S. economy using matched Census Bureau and patent data, we observe a modest departure for external R&D from perfect scaling frameworks.

JEL Classification: O31, O33, O41, L16.

Keywords: Endogenous Growth, Innovation, External, Internal, Research and Development, Patents, Citations, Scientists, Entrepreneurs.

*Comments are appreciated and can be sent to uakcigit@econ.upenn.edu and wkerr@hbs.edu. We thank Daron Acemoglu, John Haltiwanger, Sam Kortum, Dirk Krueger, Rasmus Lentz, Matt Mitchell, and many seminar participants for their insights. This research is supported by Harvard Business School, Innovation Policy and the Economy forum, Kauffman Foundation, and University of Pennsylvania. Kerr is a research associate of the Bank of Finland and thanks the Bank for hosting him during a portion of this research. Selman Erol, Kaushik Ghosh, and especially Harun Alp provided excellent research assistance on this project. The research in this paper was conducted while the authors were Special Sworn Status researchers of the US Census Bureau at the Boston Census Research Data Center (BRDC). Support for this research from NSF grant ITR-0427889 [BRDC] is gratefully acknowledged. Research results and conclusions expressed are the authors' and do not necessarily reflect the views of the Census Bureau or NSF. This paper has been screened to ensure that no confidential data are revealed.

1 Introduction

A large body of empirical research demonstrates that innovation comes in many shapes and sizes: e.g., internal vs. external, product vs. process, radical vs. normal. Similarly, we know that small and large firms can differ in their motives for pursuing innovation and the efficiency of their R&D efforts. Some accounts emphasize the many great breakthroughs of independent entrepreneurs, while others describe the financial might and longer investment horizons that large companies can afford to take towards innovation. Either way, it is clear that a Silicon Valley start-up may behave differently from the R&D laboratory of General Electric.

The goal of this paper is to embed some of this rich heterogeneity into a fully-specified endogenous growth model. Despite many advances, growth theory mostly provides frameworks that include a single type of innovation, perhaps drawn from a distribution, not the variation in types present in the data. Similarly, while models often specify a firm size distribution, the economic consequences of this distribution are typically quite limited. We seek to build a framework that allows both of these heterogeneities and links them together—an economy with firms of multiple sizes that are making different choices over the types of innovations to pursue.

The seminal model of Klette and Kortum (2004) provides the first piece for this effort. The Klette and Kortum (2004) framework allows firms to own multiple product lines, which are continually added or lost on the basis of innovation and creative destruction forces. The theoretical depictions of Klette and Kortum (2004) and the empirical quantification of Lentz and Mortensen (2008) show that this set-up exhibits many behaviors consistent with the applied micro literature (e.g., skewness of the firm size distribution, greater growth volatility of small firms). Following Lentz and Mortensen (2008), many researchers find this framework a powerful platform for applied growth theory, and we use it ourselves in Acemoglu et al. (2013) and Acemoglu et al. (2015). This framework does not, however, model multiple forms of innovation that we seek to study in the data. There are also potential roles for the firm size distribution that are not included in the model because the R&D capacities of firms are identical and scale perfectly with firm size (indeed, this resulting size independence for innovation choices provides the analytical beauty of the framework).

This study introduces into this framework new heterogeneity in the types of innovations undertaken by firms, which in turn shapes how the firm size distribution can matter for the economy. To do so, we distinguish two types of R&D that firms undertake: *external* and *internal*. Firms undertake external R&D to create new products and capture leadership in markets, while internal R&D efforts improve product lines that firms currently own. While there are other important heterogeneities across the firm size distribution, we pursue this one given the central role of innovation in endogenous growth frameworks and the many accounts of differences in innovation for large and small firms.

Our paper makes three key theoretical advances. The first is to build a growth model that incorporates multiple forms of innovation, a direct connection from firm size to choices over types of innovations, and multiple step sizes in the impact of innovations that are endogenously determined.

Our baseline framework analyzes a setting where internal R&D efforts scale up with firm size, while external R&D efforts do not. This model remains very tractable and yields analytical solutions, making stark predictions about how the R&D behavior of new entrants and small firms will differ from large firms. We show that this model can provide micro-founded explanations for observed empirical relationships like deviations from Gibrat's law of proportionate firm growth and the perceived disproportionate role of small firms for major innovations. The Klette and Kortum (2004) model focuses on external R&D, and we will consider a framework that extends the spirit of the Klette and Kortum (2004) setting to allow both internal and external innovation to scale perfectly with firm size. This comparison illuminates where our theoretical deviations occur and why.

The second theoretical contribution is to explicitly incorporate patents and patent citations into our endogenous growth framework. Using a few simple regularities from the empirical literature about patent citations, we characterize how these features would look in our economy and derive additional analytical solutions about the information embedded in patent citations. While these additions do not impact the fundamental operation of our model's economy (e.g., firms don't block rivals with patents in these models), citations are extremely important for the depth of outcomes and results that we can characterize. For example, we derive tests that employ patent citations to determine if the growth spillover effects from external innovations are larger or smaller than those from internal innovations. Moreover, we show how distributions of patent citations contain much of the information that we need to quantify external vs. internal differences in the model. This advance provides considerable scope for future work as the model can then be related to micro-empirical data with millions of observations on innovation.

Our third major contribution is to then build a generalized framework that allows an arbitrary amount of scaling for external innovation with firm size (internal innovation always scales fully). At the extremes of this generalized framework are the extended Klette and Kortum (2004) framework (perfect scaling) and our baseline model (no scaling). We quantify the model using indirect inference with Census Bureau data on all patenting firms during the 1982-1997 period. Some parts of this exercise are novel and interesting from a methodological perspective. During this period, we find that the U.S. economy exhibits moderate departures from the Klette and Kortum (2004) world, but that this attractive theoretical framework is a good approximation for many applications. In other settings, researchers may deem that the heterogeneity should not be overlooked (e.g., research for the next major drug discovery). This parallels in many respects the departures that we observed empirically from Gibrat's law of proportionate growth—they exist and are meaningful, but there are many settings where they can be deemed to be less important.

In the course of the paper we provide several empirical assessments that help inform long-standing debates about the role of small vs. large firms for innovation. For instance, we show that the relative rate of major inventions is higher in small firms. We demonstrate that these distributional differences are not due to differences in research capabilities or technologies, but are instead an outcome of R&D investment choices by firms. We report here empirical work that we use to discipline the model, especially around the identification of internal and external step sizes. Our appendix and NBER working paper, Akcigit and Kerr (2010), contains an extended empirical analysis of innovation and the firm size distribution that contains further tests regarding the scaling of innovation varieties.

Our objective is to provide new data, methods, and insights for this ongoing debate, recognizing that one paper cannot answer every part of these complex questions. Moreover, we have every reason to believe the degree of innovation scaling and the saliency of the firm size distribution may differ across times and places. The history of innovation in the United States, for example, exhibits periods that seem to bring larger or smaller firms into greater favor (e.g., Lamoreaux et al., 2011, Nicholas, 2014), and the legends of great innovation contain equal doses of Bell Laboratories and the iconic garage of Bill Hewlett and David Packard. Similarly, it is very reasonable to suspect that the scaling relationship may be very different today across Finland, Japan and India. The particular outcomes we observe for the recent United States are informative, but our real hope is that the methodology and machinery developed here will be applied in many settings to collect a broad set of evidence.

In terms of the literature, we most clearly build on the efforts of Klette and Kortum (2004), Lentz and Mortensen (2008), and Akcigit (2010) to build more insights from the empirical literature on innovation into workhorse theoretical models. More recent contributions are Lentz and Mortensen (2014) and Garcia-Macia et al. (2014). These papers in turn depend upon the long endogenous growth literature.[1] Our work on spillover benefits builds upon contributions like Spence (1984) and Griliches (1992), with Caballero and Jaffe (1993) and Eeckhout and Jovanovic (2002) being rare examples that connect patent citations to a growth model. Finally, we are deeply connected to the empirical literature on firm size and innovation that we review next as a prelude to our model.[2]

Differences in Innovation Across the Firm Size Distribution

Our work builds upon a deep empirical literature examining differences in innovation by type of firm. It is important to emphasize from the outset what our framework does *not* do. A frequent debate is whether small or large firms contribute disproportionately as the source of radical innovations; a related topic of conversation is whether small or large firms achieve a greater innovation return per dollar invested.[3] Our model would be extremely uninteresting if we obtained answers

[1] Classics include Aghion and Howitt (1992), Aghion et al. (1997), Aghion et al. (2001), Grossman and Helpman (1991), Howitt (1999), Jones (1995), Kortum (1997), and Romer (1986, 1990). Barro and Sala-i Martin (1995) and Acemoglu (2008) provide full reviews.

[2] Our work likewise relates to the economics literatures on innovation and industry structure and evolution. Examples include Acemoglu and Cao (2015), Acemoglu and Akcigit (2012), Arkolakis (2011), Bloom et al. (2013), Cabral and Mata (2003), Cai (2010), Cohen (1995), Dunne et al. (1988), Duranton (2007), Gans et al. (2002), Gilbert and Newbery (1982), Hausman et al. (1984), Hopenhayn (1992), Hopenhayn et al. (2006), Jovanovic (1982), Jovanovic and MacDonald (1994), Kerr (2010), Klepper and Graddy (1990), Lerner (1997), Luttmer (2007, 2011), Reinganum (1983), and Rosen (1991).

[3] For example, Acemoglu et al. (2015), Acs and Audretsch (1987, 1988, 1991), Baumol (2002), Kerr et al. (2014), Kueng, Yang, and Hong (2014), Nelson and Winter (1982), Peretto (1998), Rausch (2010), Rosen (1991), Samila and Sorenson (2011), Thomke (2003), and Zucker et al. (1998). Of the efforts to quantify these claims, the best known is the Kortum and Lerner (2000) finding that venture capital dollars invested in small start-ups are three times more

to these types of questions by giving some firms the exogenous power to produce certain types of innovations not available to others (e.g., that the innovation distribution for small firms showed greater variance than the distribution of larger firms). Instead, we want to trace out why large and small firms might invest at different rates in the same set of potential innovation approaches, with differences in the size of realized advances being an outcome rather than an assumption.

We focus on internal vs. external innovation as it aligns with many important insights about the innovation behavior of small and large companies. It will also be the most direct heterogeneity that we can measure with the data. At an extreme, external vs. internal differences must exist. Entering entrepreneurs, in our model and in the real world, do not have products to improve upon, and so by definition we start to expose differences in potential behavior. What follows are collected insights as to why this difference is pervasive rather than confined to the extremes, usually emphasizing a greater internal focus for large firms.

One rationale for why large firms might invest more in internal improvements is that they can derive a better return from these investments than small firms can. In situations where innovations are useful for enhancing a company's operations but will be otherwise hard to protect/sell, large companies achieve a greater return for the same investment due to their larger base of operations. These incentives have been most frequently discussed for product vs. process R&D investments (e.g., Klepper, 1996), and Cohen and Klepper (1996) show process R&D is more tightly linked to firm size than product R&D. While these patterns are consistent with internal R&D scaling more directly with firm size, at least one counter example exists. Basic R&D is also more likely to be conducted by large firms due to the fixed costs of basic R&D laboratories and the ability to realize resulting discoveries across a range of products. To the extent that basic R&D also provides serendipitous advances that aid entry into new industries outside of the firm's current span, larger companies garner more external innovation.

An additional class of explanations for why large companies may pursue proportionately less external innovation relates to organizational frictions and managerial capabilities. In one perspective, following on the Lucas (1978) span-of-control model, this is efficient. If there are limits to the number of operations that the world's best managers can effectively guide, then large companies might endogenously invest more in improving their existing products vs. conquering new domains. These limits to optimal firm size effectively give a comparative advantage to others for pursuing the acquisition of new lines.[4] Related, the model of Hellmann and Perotti (2011) emphasizes the prowess of large companies for innovations that require significant circulation of innovative ideas over internal resources to complete, while clusters of smaller companies have advantages in other settings. Gromb and Scharfstein (2002) model related trade-offs.

On the other hand, many accounts in the management literature reach the same conclusion but

potent for generating patented innovations than corporate R&D expenditures.

[4] Akcigit et al. (2015) introduces a span-of-control limitation for firms into the Klette and Kortum (2004) framework to study the firm dynamics in the Indian manufacturing sector. Their model emphasizes how these managerial conditions create limits to the scaling of firms and their pursuit of new product lines, connecting to the empirical work of Hsieh and Klenow (2014).

stress inefficient origins. These accounts suggest that organizational rigidities and pre-occupation with serving existing customers stifle external innovations that large companies want to pursue (e.g., Christensen, 1997, Clark and Henderson, 1993, Henderson, 1990, March, 1991). Galasso and Simcoe (2011) identify how CEO personality traits shape innovation investments, and Lerner (2012) further reviews the recent literature on the advantages and liabilities of large companies for pursuing new innovation areas compared to start-ups (e.g., compensation constraints).

There is also evidence that external environments shape innovation incentives for large companies. Financial markets provide a well-studied example. Using plausibly exogenous variation in going public decisions, Bernstein (2015) finds that being a publicly listed firm reduces the novelty of a firm's innovations by up to 40% and shifts work towards more conventional and internal projects, while perhaps offering some compensation in funds for acquisitions. Lerner et al. (2011) reach similar conclusions when examining the impact of private equity firms on the innovation rates of the firms that they remove from public markets. Other studies find conglomerate firms frequently trade at a discount, and that managers often reduce R&D budgets to meet short-term return targets. Thus, while deep capital markets may provide valuable resources to public companies, they appear to create environments less attractive for external innovation.[5]

This brief literature discussion highlights why the internal vs. external distinction is likely to be important. Several data sources are consistent with this observation:[6]

- Using the 2008 Business R&D and Innovation Survey, we observe a -0.16 correlation between firm size and the share of R&D that the firm reports is directed towards business areas and products where the company does not have existing revenues. Similar negative correlations are found for questions about the share of firm R&D being directed to technologies new to markets.

- Using the 1979-1989 NSF R&D Surveys that recorded product vs. process R&D expenditures, we observe a 0.22 correlation between firm size and the share of R&D that the firm reports is process oriented. This accords with Cohen and Klepper (1996), and we find similar results for indicator variables about the firm conducting any process-focused R&D.

- Using the citations that firms make on the patents they file, we observe a 0.11 correlation between firm size and the share of backward citations that are made to a firm's own prior work. Firms with larger past patent portfolios are mechanically more likely to self cite, and

[5]Differences beyond financial markets also exist. Agrawal et al. (2010) consider how large companies may be located in more isolated cities that limit the diversity of external ideas that they receive and can build upon. Some industries are also characterized by a market for ideas (e.g., Gans et al. (2002)) that shifts the organization of innovation for external work. Finally, policies with firm-size-dependent components like labor regulations may make external innovation less attractive for large companies to the extent that policies make the labor adjustments associated with risky activities more costly for larger employers.

[6]All reported correlations measure firm size through log employment and are statistically significant at a 5% level. The correlations are taken over reported data in each survey, and some of these sources have incomplete coverage for small R&D producers, as described in our working paper. These sample constraints likely weaken the observed correlations to firm size.

the appendix C1 reports Monte Carlo simulations that measure the expected likelihood of self citations given the technology and years that a firm cites in their work. Larger firms are more likely to show abnormal rates of self citations compared to these counterfactuals, with the correlation to firm size of being out of the simulated 95th bound being 0.23.

These correlations, while containing their own advantages and liabilities, each point towards a consistent picture of heterogeneity in R&D behavior by firm size. An advantage of our model is its capacity to place these data pieces into context and use indirect inference for more general statements.

As a final observation, we emphasize that our focus is solely on innovative firms, which in our empirical exercises will be the universe of U.S. firms that file patents. There are many, many small firms that are not engaged in innovation and have no expectation to ever do so or to grow larger than their current size (Hurst and Pugsley, 2011). This is often connected to non-pecuniary motivations for starting a business (e.g., to be one's own boss). Our study does not consider this mass of small businesses, and our large-small depictions should be viewed only in terms of innovative firms. An important insight from work like Hurst and Pugsley (2011) is the degree to which the "Main Street" small business sector is very different from innovative start-ups and that this imprinting is almost always present at founding. Our concentration is on innovative firms connected to technological advancement, and the fascinating questions of how to work in the many Main Street small businesses into growth models, or to model the quite rare transitions of existing non-innovation firms into innovation, will need to be saved for later.

2 Baseline Theoretical Framework

We begin with a baseline model that incorporates the empirical regularity that external R&D does not scale as fast as internal R&D with firm size. Our goal is to study the implications of this heterogeneity on the R&D, innovation and growth dynamics of firms. To allow for analytical solutions and to build intuition, we first consider a stark environment where external R&D does not scale at all with firm size. We then generalize the theoretical framework to allow scaling of external R&D, with this baseline model and Klette and Kortum (2004) being extremes of the general framework. On top of this general framework, we also overlay patent citation behavior so that we can quantitatively estimate the scaling parameter. The appendix contains proofs of propositions.

2.1 Preferences and Final Good Technology

Consider the following continuous time economy. The world admits a representative household with a logarithmic utility function

$$U = \int_0^\infty \exp\left(-\rho t\right) \ln C\left(t\right) dt. \tag{1}$$

$C(t)$ is consumption at time t, and $\rho > 0$ is the discount rate. The household is populated by a continuum of individuals with measure one. Each member is endowed with one unit of labor that is supplied inelastically.

Individuals consume a unique final good $Y(t)$, which is also used for R&D as discussed below. The final good is produced by labor and a continuum of intermediate goods $j \in [0,1]$ with the production technology

$$Y(t) = \frac{L^{\beta}(t)}{1-\beta} \int_0^1 q_j^{\beta}(t)\, k_j^{1-\beta}(t)\, dj. \tag{2}$$

In this specification, $k_j(t)$ is the quantity of intermediate good j, and $q_j(t)$ is its quality. We normalize the price of the final good Y to be one in every period without loss of generality. The final good is produced competitively with input prices taken as given. Henceforth, the time index t will be suppressed when it causes no confusion.

There is a set of firms that are producing intermediate goods and their measure, $F \in (0,1)$, will be determined in equilibrium. Each intermediate good j is owned by a firm f. A firm is characterized by the collection of its product lines $\mathcal{J}_f = \{j : j \text{ is owned by firm } f\}$. Similarly we denote the product (quality) portfolio of firm f by a *multiset* $\mathbf{q}_f = \{q_j : j \in \mathcal{J}_f\}$ and denote the cardinality by n_f.[7] Each intermediate good $j \in [0,1]$ is produced with a linear technology

$$k_j = \bar{q} l_j, \tag{3}$$

where l_j is the labor input and $\bar{q} \equiv \int_0^1 q_j dj$ is the average quality in the economy. This linear specification has two implications. First, intermediate goods have the same marginal cost $w/\bar{q} > 0$, where w is the wage rate in terms of the final good. Second, the marginal product of labor in the intermediate good sector grows at the same rate as in the final good sector, generating an invariant labor allocation across sectors in steady state.

In addition to the variable cost, production requires also a fixed cost of operation $\Phi\bar{q}$ at the firm level in terms of the final good. As it will transpire later, this fixed cost will be used to avoid any non-linearities in the firm's value function.[8]

Individuals work in two capacities: final good production (L) and intermediate good production ($\tilde{L}$). In each period, the labor market has to satisfy the constraint

$$L + \tilde{L} \leq 1. \tag{4}$$

R is the total R&D spending, K is the total fixed cost paid by firms, and therefore the resource constraint of the economy is $Y = C + R + K$.

[7] A multiset is a generalization of a set which can contain more than one instances of the same member.

[8] See Proposition 1 and the text above it for details.

2.2 Research and Development

The last innovator in each product line owns the leading patent and has monopolist pricing power until being replaced by another firm. Intermediate producers have profit incentives to improve the technologies for their existing products, thereby increasing associated quality. In addition, both incumbents and potential entrants have incentives to add new products to their portfolios through R&D competition. We now describe the innovation types.

Internal R&D Incumbent firms undertake *internal R&D* (or innovation) to improve their existing products. To improve an existing product $j \in \mathcal{J}_f$, firm f spends

$$R_z(z_j, q_j) = \hat{\chi} z_j^{\hat{\psi}} q_j \tag{5}$$

units of the final good, where $\hat{\chi} > 0$ and $\hat{\psi} > 1$. Internal innovations are realized with the instantaneous Poisson flow rate of $z_j \geq 0$. Cost (5) is proportional to the quality of the good that the firm is improving. First, this implies that a more-advanced technology has higher R&D costs. Second, as will be shown in the next section, equilibrium returns to internal innovations are linear in q_j. Therefore, the linear effects in return and cost cancel out and yield an internal innovation effort that is independent of the quality of the product line. When internal R&D is successful, the current quality improves by a multiplicative factor $\lambda > 0$ such that $q_j(t + \Delta t) = (1 + \lambda) q_j(t)$.

External R&D *External R&D* (or innovation) is undertaken by incumbents and potential new entrants to obtain technology leadership over products that they do not currently own. A firm produces a flow rate x by paying R_x in terms of the final good according to the following production function:

$$x = \left[\frac{R_x}{\tilde{\chi}\bar{q}}\right]^{\frac{1}{\tilde{\psi}}} \mathbf{1}_{n>0}$$

where $\mathbf{1}_{n>0}$ is an indicator function. This specification implies that past innovation, i.e. $n > 0$, affords firms capacities to innovate in the future. This structure is in the same spirit as the Klette and Kortum (2004) model that assumes a Cobb-Douglas functional form: $x = R^{\frac{1}{\tilde{\psi}}} n^{1-\frac{1}{\tilde{\psi}}}$. For now, we shut down the dependence on n at the intensive margin to prevent any scaling and just keep the dependence on the extensive margin via the indicator function. The resulting cost function for a firm with $n > 0$ is

$$R_x(x, \bar{q}) = \tilde{\chi} x^{\tilde{\psi}} \bar{q}, \tag{6}$$

where $\tilde{\chi} > 0$ and $\tilde{\psi} > 1$. Cost (6) is proportional to the average quality level $\bar{q}$ in the economy, which again removes the dependence of innovation efforts on average quality since the returns to external innovations will be proportional to $\bar{q}$ and ensures that the R&D spending is a constant fraction of the total output Y.

External R&D efforts are undirected in the sense that resulting innovations are realized in any product line $j \in [0, 1]$ with equal probability. This model structure has two main implications.

First, firms do not innovate over their own product lines through external R&D since this event has zero probability. Second, there is no strategic interaction among firms. In addition to stochastic arrival rates, the sizes of realized quality improvements are randomly determined:

(i) With probability $\theta \in (0, 1)$, the innovation is a *major advance* that substantially shifts forward the latest quality level by a size $\eta\bar{q}$ such that $q_j(t + \Delta t) = q_j(t) + \eta\bar{q}(t)$. This generates a new technology cluster with an associated wave of subsequent follow-on innovations. Prominent examples include the transistor and mapping the human genome, but the step functions need not be so profound. The conceptual construct is that these major advances define a wave of innovation and product development until another major advance starts a new wave.

(ii) With probability $1 - \theta$, the innovation is a *follow-up improvement* to the current technology level of the product line that does not generate a new technology cluster. The size of the follow-up improvement declines with the number of follow-up inventions since the last major advancement. If the last major innovation in product line j occurred k_j innovations ago, the new step size is $s_j\bar{q}$, where $s_j = \eta\alpha^{k_j}$ with $\alpha \in (0, 1)$.

Technology Clusters and Evolution The economy-wide arrival rate of new products, denoted by τ, is endogenously determined by external R&D efforts of incumbents and potential entrants and is characterized in detail below. With τ determined, the probabilistic evolution of the quality level q_j after a short interval Δt is

$$q_j(t + \Delta t) = q_j(t) + \begin{cases} \eta\bar{q}(t) & \text{with probability} \quad \theta\tau\Delta t \\ \eta\alpha^{k_j}\bar{q}(t) & \text{with probability} \quad (1-\theta)\tau\Delta t \\ \lambda q_j(t) & \text{with probability} \quad z_j\Delta t \\ 0 & \text{with probability} \quad 1 - z_j\Delta t - \tau\Delta t \end{cases}$$

The first line represents a major advance that results from external R&D with probability θ. The second line represents a follow-up innovation that results from external R&D with probability $1-\theta$. The third line shows an internal improvement of size λ by the current owner of product line j through internal R&D. The final line represents the case where no quality improvement is realized during Δt, which results in stagnant technology quality.

The following example illustrates a possible evolution of innovations in a random product line:

Example 1

$$\Big| \underbrace{\frac{\eta}{P_{1,f_1}} \frac{\eta\alpha}{P_{2,f_2}} \frac{\eta\alpha^2}{P_{3,f_3}} \frac{\lambda}{P_{4,f_3}} \frac{\lambda}{P_{5,f_3}} \frac{\eta\alpha^3}{P_{6,f_4}}}_{\text{Tech Cluster 1}} \Big| \underbrace{\frac{\eta}{P_{7,f_5}} \frac{\lambda}{P_{8,f_5}} \frac{\eta\alpha}{P_{9,f_6}}}_{\text{Tech Cluster 2}} \Big| \underbrace{\frac{\eta}{P_{10,f_7}} \frac{\eta\alpha}{P_{11,f_8}} \frac{\eta\alpha^2}{P_{12,f_9}} \ldots}_{\text{Tech Cluster 3}}$$

An example of a sequence of innovations in a product line

Here, $P_{m,f}$ denotes that the mth patent is obtained by firm f. The example starts with a major innovation that opens a new technology cluster by firm f_1. Firms f_2 and f_3 then produce follow-up external innovations. Firm f_3 further improves its own product twice. Firm f_4 then produces a further follow-up external innovation. Next, this technology cluster is replaced by a new leading innovation by firm f_5 which is patented as P_7. The second cluster is then replaced by another leading innovation by firm f_7. This new cluster is further improved by patents 11 and 12, and so on.

We later analytically solve for an expected step size $\bar{s}$ from external innovations. For now, it is important to note that this theoretical structure does not depend upon $\bar{s}$ being greater or smaller than λ, and in fact this comparison may differ substantially depending upon the country and time period studied. The baseline model framework is very general with respect to the relative sizes of internal versus external improvements.

2.3 Entry and Exit

As in Klette and Kortum (2004), a mass of entrants invest in R&D in order to become intermediate producers upon a successful innovation. Entrants choose an innovation flow rate $x_e > 0$ with an R&D cost $C_e(x_e, \bar{q}) = x_e \nu \bar{q}$ in terms of the final good, where $\nu > 0$ is a constant scale parameter. The value V_0 of being an outside entrepreneur is the expected value from innovating successfully and entering the market. This value is determined according to

$$rV_0 - \dot{V}_0 = \max_{x_e} \left\{ x_e \left[\mathbb{E}_j V\left(\{q_j + s_j \bar{q}\}\right) - V_0 \right] - \nu x_e \bar{q} \right\}, \tag{7}$$

where $V(\{q_j\})$ denotes the value of a firm that owns a single product line with quality q_j and $\dot{V}_0 \equiv \partial V_0 / \partial t$ denotes the partial derivative of the outside value with respect to time. The expected value $\mathbb{E}_j V(\{q_j + s_j \bar{q}\})$ of a new innovation is an expectation over both quality level q_j and innovation size s_j. When there is positive entry, the equilibrium is such that

$$\mathbb{E}_j V\left(\{q_j + s_j \bar{q}\}\right) = \nu \bar{q}. \tag{8}$$

Incumbent firms produce intermediate inputs and invest in R&D. As a result, firms simultaneously expand into new product lines and lose some of their current product lines to other firms in the economy through competition. Each product line faces the same aggregate endogenous creative destruction rate τ. A firm that loses all product lines to competitors exits the economy.

2.4 Equilibrium

We now characterize the Markov Perfect Equilibria of the economy that make strategies a function of payoff relevant states only. We focus on the steady state in which aggregate variables (Y, C, R, K, w, $\bar{q}$) grow at the constant rate g.

2.4.1 Production

The standard maximization problem of the representative household yields the Euler equation

$$\frac{\dot{Y}}{Y} = \frac{\dot{C}}{C} = r - \rho. \tag{9}$$

The maximization problem of the final goods producer generates the inverse demand $p_j = L^{\beta} q_j^{\beta} k_j^{-\beta}$, $\forall j \in [0,1]$. The constant marginal cost of producing each intermediate variety is $w/\bar{q}$.

The profit maximization problem of the monopolist j is thus,

$$\pi(q_j) = \max_{k_j \geq 0} \left\{ L^{\beta} q_j^{\beta} k_j^{1-\beta} - \frac{w}{\bar{q}} k_j \right\} \; \forall j \in [0,1]. \tag{10}$$

The first order condition (10) yields an optimal quantity and price for intermediate good j

$$k_j = \left[\frac{(1-\beta)\,\bar{q}}{w} \right]^{\frac{1}{\beta}} L q_j \text{ and } p_j = \frac{w}{(1-\beta)\,\bar{q}}. \tag{11}$$

The realized price is a constant markup over the marginal cost and is independent of the individual product quality. Thus, the profit for each active good is $\pi(q_j) = \pi q_j$, where $\pi \equiv L(\bar{q}/w)^{\frac{1-\beta}{\beta}} (1-\beta)^{\frac{1-\beta}{\beta}} \beta$. In order to avoid the case of limit pricing and maintain a simple model, we adopt the following stage-game assumption.

Assumption 1 (Monopoly pricing) *In a given product line j, the current incumbent and any former incumbents in the same line (with lower quality than the current incumbent) enter a two-stage price-bidding game. In the first stage, each firm pays a fee of ϵ which is arbitrarily close to 0. In the second stage, all firms that paid the fee announce their prices.*

Under Assumption 1, only the leader pays the fee and enters the second stage since other firms can never recover their fee in the second stage. Since the leader is the only firm bidding a price, the leader will always operate with monopoly pricing, as in Aghion and Howitt (1992).

The maximization in the final goods sector, together with (11), implies a wage rate

$$w = \tilde{\beta}\bar{q}, \tag{12}$$

where $\tilde{\beta} \equiv \beta^{\beta} [1-\beta]^{1-2\beta}$. Incorporating the equilibrium wage rate, the constant part of the equilibrium profit simplifies to

$$\pi = L(1-\beta)\,\tilde{\beta}. \tag{13}$$

Note that, using the equilibrium quantity (11) and the wage rate (12), the aggregate output can now be expressed as as a linear function of production workers L and the average quality $\bar{q}$ such that

$$Y = \frac{[1-\beta]^{1-2\beta}}{\beta^{1-\beta}} \bar{q} L. \tag{14}$$

Equations (3), (4) , (11), and (12) determine the final good workers as

$$L = \frac{\beta}{(1-\beta)^2 + \beta}. \tag{15}$$

2.4.2 Invariant Step-size Distribution and $\bar{s}$

We next compute the invariant step-size distribution $\Psi(s)$ that determines the expected innovation size from external innovations $\bar{s}$. Let Ψ_k denote the equilibrium share of product lines with $k \in \mathbb{N}_0$ subsequent follow-up innovations such that $s_j = \eta\alpha^{k_j}$. A steady state equilibrium requires a stable innovation size distribution. Thus, while the stochastic nature of innovation moves individual products up and down the k distribution, the overall share of products at each level k is stable. This stability requires equal inflows and outflows of products from each size level, resulting in the flow equations

$$\begin{array}{rcll} \textsc{Outflow} & & \textsc{Inflow} & \\ \Psi_0 \tau (1-\theta) & = & (1-\Psi_0)\tau\theta & (16) \\ \Psi_k \tau & = & \Psi_{k-1}\tau(1-\theta) \text{ for } k \geq 1. & (17) \end{array}$$

The first line governs inflows and outflows among product lines where major innovations have just occurred. Outflows happen due to follow-up innovations at the rate $\tau(1-\theta)$, while inflows happen due to new leading innovations being realized at rate $\tau\theta$ throughout the innovation size distribution. Internal R&D within firms does not influence these k distributions. A similar reasoning governs the share of product lines with $k \geq 1$ consecutive follow-up innovations in (17). As a result, flow equations (16) and (17) generate the invariant distribution

$$\Psi_k = \theta(1-\theta)^k \text{ for } k \geq 0, \tag{18}$$

which yields the expected innovation size from external R&D:

$$\bar{s} = \mathbb{E}(s_j) = \sum_{k=0}^{\infty} \Psi_k \eta \alpha^k = \frac{\theta\eta}{1-(1-\theta)\alpha}. \tag{19}$$

This expected size is naturally increasing in the probability of a major innovation θ, the realized size of major innovations η, and for lower decay rates in innovation quality within a technology cluster (i.e., higher α).

2.4.3 Research and Development by Incumbents

The value functions of firms determine R&D choices. For simplicity we drop the firm subscript f from the firm variables when it causes no confusion. Consider a firm with a product portfolio $\mathbf{q}$ which serves as the state variable in the firm's problem. The firm takes the values of (r, τ, g) as given and chooses the optimal R&D efforts x and z_j for every $j \in \mathcal{J}$ to maximize the following value function:[9]

$$rV(\mathbf{q}) - \dot{V}(\mathbf{q}) = \max_{\substack{x\in[0,\bar{x}],\\ \{z_j\in[0,\bar{z}]\}_{\mathcal{J}}}} \left\{ \begin{array}{c} \sum_{q_j\in\mathbf{q}} \left[\begin{array}{c} \pi q_j - \hat{\chi} z_j^{\hat{\psi}} q_j \\ +z_j \left[V\left(\mathbf{q} \setminus_- \{q_j\} \cup_+ \{q_j(1+\lambda)\}\right) - V(\mathbf{q}) \right] \\ +\tau \left[V\left(\mathbf{q} \setminus_- \{q_j\}\right) - V(\mathbf{q}) \right] \end{array} \right] \\ +x\left[\mathbb{E}_j V\left(\mathbf{q} \cup_+ \{q_j + \bar{q}s_j\}\right) - V(\mathbf{q}) \right] \\ -\tilde{\chi} x^{\tilde{\psi}} \bar{q} - \Phi \bar{q} \end{array} \right\}. \tag{20}$$

The first line on the right hand side represents operating profits over currently held product lines minus R&D costs. The second line is the change in firm value after internal improvements to currently held products. $V(\mathbf{q}\setminus_- \{q_j\} \cup_+ \{q_j(1+\lambda)\})$ denotes the firm value after improving one of the firm's existing products by size λ. These terms are multiplied by the Poisson arrival rate z_j as the success of internal R&D is stochastic. Firms choose innovation effort for each line separately. The third line shows the change in firm value due to losing its product lines through creative destruction τ. $V(\mathbf{q}\setminus_- \{q_j\})$ denotes firm value after losing product that had quality q_j.

The fourth line is the change in firm value after a successful external innovation that garners a new product line. $V(\mathbf{q} \cup_+ \{q_j + \bar{q}s_j\})$ denotes equilibrium firm value after a successful external innovation of size s_j that adds a new product into the firm's portfolio. With probability θ, external R&D generates a major advance. With probability $(1-\theta)$, a follow-up advance occurs. In the case of a major innovation, the step size is η. For follow-up advances, the innovation size is $s_j = \eta\alpha^{k_j}$ where $k_j > 0$. These terms are multiplied by the Poisson arrival rate x as the success of external R&D is stochastic too. The final line represents external R&D costs and fixed costs. The $-\dot{V}(\mathbf{q})$ term on the left hand side of equation (20) represents change in firm value without any material events for the focal firm due to economy-wide growth (i.e., $\bar{q}$ changes).

The aggregate growth rate is determined by the frequency of innovations coming from creative destruction τ, consisting of new entry x_e and external innovations x by incumbents; the frequency of internal innovations z; and their relevant innovation sizes as described in the following lemma.

Lemma 1 *Let the equilibrium R&D efforts be given by (τ^*, z^*). The steady state growth rate of the aggregate variables in the economy is*

$$g = \tau^* \bar{s} + z^* \lambda. \tag{21}$$

Now we are ready to solve for the equilibrium value function. One technical detail needs particular attention. Our goal in this benchmark model is to generate new intuitions while preserving tractability. The Klette and Kortum (2004) model is very tractable since everything scales perfectly in the number of product lines of the firms; this includes the profits collected by the firm and

[9]We do not index the portfolio or R&D efforts by f as $\mathbf{q}_f$, x_f and $z_{j,f}$ to simplify notation. $\cup_+$ indicates the multiset union operator such that $\{a,b\} \cup_+ \{b\} = \{a,b,b\}$. Similarly $\setminus_-$ indicates the multiset difference operator such that $\{a,b,b\} \setminus_- \{b\} = \{a,b\}$.

the *franchise* value, which is an option value for external innovation arising due to the fact that more product lines makes the firm more innovative via the Cobb-Douglas R&D technology. In our baseline model, profits also scale perfectly, yet the franchise value is constant across all firms since the R&D technology depends on having positive product lines only at the extensive margin but not on the intensive margin. This introduces a non-linearity to the firm value function. To generate a value function that scales perfectly with the number of product lines as in the Klette and Kortum (2004) model, we assume that the fixed cost of operation is equal to the franchise value as follows.[10]

Assumption 2 (Perfectly-scaling value function) *The value of fixed cost of operation satisfies*

$$\Phi = \left[\frac{\nu}{\tilde{\psi}\tilde{\chi}}\right]^{\frac{\tilde{\psi}}{\tilde{\psi}-1}} \tilde{\chi}\left(\tilde{\psi}-1\right).$$

The next proposition shows that the value function (20) and its components can be expressed in a very tractable form. We assume for now that there is positive entry and later impose a parameter restriction that is sufficient to verify this condition.

Proposition 1 *Under assumptions 1 and 2 and when there is positive entry $x_e > 0$, the value function (20) of a firm with a set of product lines $\mathbf{q}$ can be expressed as $V(\mathbf{q}) = A \sum_{q_j \in \mathbf{q}} q_j$ where A (the value of holding a product line) is*

$$A = \frac{\nu}{1+\bar{s}}. \tag{22}$$

Moreover, the optimal R&D decisions are given by

$$z_j = \left[\frac{\lambda\nu}{(1+\bar{s})\,\hat{\psi}\hat{\chi}}\right]^{\frac{1}{\hat{\psi}-1}} \text{ and } x = \left[\frac{\nu}{\tilde{\psi}\tilde{\chi}}\right]^{\frac{1}{\tilde{\psi}-1}}, \tag{23}$$

and the aggregate creative destruction rate is

$$\tau = \frac{1}{(1+\bar{s})}\left[\frac{\pi}{A} - \left[\frac{\lambda}{\hat{\psi}\hat{\chi}}\right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} A^{\frac{1}{\hat{\psi}-1}}\hat{\chi} - \rho\right]. \tag{24}$$

This proposition shows that the innovation efforts of incumbents, both internal and external, are positively related to the entry cost. Higher entry costs lower entry rates and thus provide longer expected durations and profits from owning product lines. Moreover, both internal and external R&D efforts decline in their own cost scale parameters.

Importantly, internal innovation is increasing in its own step size λ due to higher marginal return to successful internal improvements, but internal investments are decreasing in the average

[10] The equality simplifies the math for the rest of the baseline model. These technical conditions related to fixed costs are not important for our general framework, and thus fixed costs are set equal to zero in later sections.

step size of external innovation $\bar{s}$, since larger $\bar{s}$ encourages more creative destruction that lowers the expected duration of monopoly power the firm has on the product line. By contrast, step sizes do not show up in the equilibrium external innovation rate since a bigger step size $\bar{s}$ both encourages effort (due to higher return) and discourages it (due to higher entry); these two opposing effects cancel out.

The aggregate creative destruction rate is the sum of average external innovation effort by each incumbent, Fx, and the realized entry rate x_e,

$$\tau = Fx + x_e. \tag{25}$$

To pin down the actual entry rate, we solve for the equilibrium measure of firms F. To achieve this, we first characterize the invariant distribution of the number of products. This distribution is the main proxy for the firm size distribution in Klette and Kortum (2004). Let μ_n denote the equilibrium share of the incumbent firms that own n product lines such that $\Sigma_{n=1}^{\infty}\mu_n = 1$. The invariant distribution again depends upon the following flow equations:

$$\begin{array}{lrl} \textsc{State}: & \textsc{Inflow} & \textsc{Outflow} \\ n=0: & F\mu_1\tau = & x_e \\ n=1: & F\mu_2 2\tau + x_e = & F\mu_1(x+\tau) \\ n\geq 2: & F\mu_{n+1}(n+1)\tau + F\mu_{n-1}x = & F\mu_n(x+n\tau) \end{array} \tag{26}$$

The first line characterizes outside entrepreneurs ($n = 0$). Inflows to outside entrepreneurs happen when firms with one product are destroyed, and outflows occur when outside entrepreneurs successfully develop a new product at rate x_e. Similarly, the second line considers inflows and outflows of firms with one product, and the third line considers n-product firms. The next proposition provides the explicit form solution of the invariant product number distribution.

Proposition 2 *The invariant distribution μ_n is equal to*

$$\mu_n = \frac{x_e}{Fx}\left(\frac{x}{\tau}\right)^n \frac{1}{n!} \textit{ for } n \geq 1. \tag{27}$$

Since (27) is a probability distribution, it must be that $\Sigma_{n=1}^{\infty}\mu_n = 1$, which implies $\frac{Fx}{x_e} = e^{\frac{x}{\tau}} - 1$. This condition and (25) deliver the entry rate as

$$x_e = \tau e^{-\frac{x}{\tau}} \text{ and } F = \frac{\tau}{x}\left(1 - e^{-\frac{x}{\tau}}\right). \tag{28}$$

The entry rate is a fraction of the aggregate creative destruction rate. In order to ensure an equilibrium with positive aggregate creative destruction and entry, we make the following assumption.

Assumption 3 (Positive Entry) *The parameters of the model are such that*

$$\pi > \left[\frac{\lambda}{\hat{\psi}\hat{\chi}}\right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} \left[\frac{\nu}{1+\bar{s}}\right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} \hat{\chi} + \frac{\nu\rho}{1+\bar{s}}.$$

This assumption is very easy to satisfy. For any given positive profit, there is always a low enough entry cost ν such that an equilibrium with positive entry exists.

The total R&D effort of the economy is

$$R = \hat{\chi}\left[\frac{\lambda\nu}{(1+\bar{s})\,\hat{\psi}\hat{\chi}}\right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} \bar{q} + F\tilde{\chi}\left[\frac{\nu}{\tilde{\psi}\tilde{\chi}}\right]^{\frac{\tilde{\psi}}{\tilde{\psi}-1}} \bar{q} + \nu\tau e^{-\frac{x}{\tau}}\bar{q}, \tag{29}$$

and the total fixed cost is

$$K = F\Phi\bar{q}. \tag{30}$$

Combining (14) and (15) delivers the equilibrium output level,

$$Y = \frac{[1-\beta]^{1-2\beta}\,\beta^{\beta}}{(1-\beta)^2+\beta}\bar{q}. \tag{31}$$

From this, consumption is determined through the resource constraint as

$$C = Y - K - R. \tag{32}$$

We end this section by summarizing the equilibrium.

Definition 1 (Balanced Growth Path Equilibrium) *A balanced growth path equilibrium of this economy consists of the following tuple for every t, $j \in [0,1]$, $\bar{q}$, and q_j : k_j^*, p_j^*, w^*, L^*, $\tilde{L}^*$, x^*, z_j^*, τ^*, x_e^*, F^*, R^*, K^*, Y^*, C^*, g^*, Ψ_n^*, μ_n^*, r^*, such that: (i) k_j^* and p_j^* satisfy (11); (ii) wage rate w^* satisfies (12); (iii) measure of final good production workers L^* satisfies (15) and $\tilde{L}^*$ is simply $1-L^*$; (iv) external (x^*) and internal (z_j^*) innovation flows are equal to (23); (v) aggregate creative destruction τ^* satisfies (24); (vi) entry flow x_e^* and measure of incumbent firms F^* satisfy (28); (vii) total R&D spending R^* satisfies (29); (viii) total amount of fixed cost expenses K^* satisfies (30); (ix) aggregate output Y^* satisfies (31); (x) aggregate consumption C^* satisfies (32); (xi) steady state growth rate g^* satisfies (21); (xii) the invariant distribution of innovation sizes Ψ_n^* satisfies (18); (xiv) the invariant distribution of number of products μ_n^* satisfies (27); and (xv) the interest rate satisfies the Euler equation (9).*

2.5 Central Theoretical Results

The following propositions characterize the firm growth, R&D, and innovation dynamics of the model. In our model, the ideal proxy for firm size is the total quality $Q = \sum_{q_j \in \mathbf{q}} q_j$. This is because firm sales, profits, and production workers are all proportional to Q.[11] Firm size also closely relates to the number of product lines, which we discuss in Section 6.3. Therefore, we also use n_f to proxy for firm size in propositions when convenient.

Proposition 3 *Let $\mathcal{G}(Q) \equiv \mathbb{E}\left(\dot{Q}/Q\right)$ be the average growth rate of a firm with total quality Q. Then $\mathcal{G}(Q)$, in equilibrium, is given by*

$$\mathcal{G}(Q) = \frac{x(1+\bar{s})\bar{q}}{Q} + z\lambda - \tau.$$

$\mathcal{G}(Q)$ is a strictly decreasing function.

This result suggests that small firms grow faster than large firms. This micro-founded departure from Gibrat's law of proportionate growth occurs due to the lack of scaling of external innovation efforts. As a result, the growth coming from internal innovation is the same on average across different firm sizes ($z\lambda$), whereas the contribution of external R&D to firm growth gets smaller as firm size increases (the first ratio in $\mathcal{G}(Q)$). Combining these effects, overall firm growth declines with firm size.

Proposition 4 *Let $\mathcal{R}(Q) \equiv R\&D/Sales$ be the firm R&D intensity of a firm with total quality Q. Then $\mathcal{R}(Q)$, in equilibrium, is given by*

$$\mathcal{R}(Q) = \frac{\beta c_x(x)\bar{q}}{\pi Q} + \frac{\beta c_z(z)}{\pi}.$$

$\mathcal{R}(Q)$ is a strictly decreasing function.

This result suggests that small innovative firms have a greater R&D intensity than large firms. Similar to the previous proposition, the intuition is that total internal R&D effort is proportionate to the number of product lines of the firm. On the other hand, external R&D efforts do not scale with number of product lines, which results in a declining R&D intensity for larger firms. In other words, adding additional product lines continually adds more R&D efforts but further dilutes the external R&D effects with respect to intensity measures.

Up until this point, our model does not require taking a stance on the relative sizes of internal vs. external innovations. With some structure added that is consistent with our later empirical results, the model also makes predictions about the innovation size distribution and the relative frequency of firms by innovation size.

Proposition 5 *Let a major innovation be defined as an innovation with a step size larger than a certain threshold $s_k \geq s_{\hat{k}}$ for some $\hat{k} \in \mathbb{Z}_+$ and $s_{\hat{k}} > \lambda$. Moreover, let $\mathcal{M}(n)$ be the probability of*

[11] $Sales = \sum_{q_j \in \mathbf{q}_f} p(q_j)\, k(q_j) = [(1-\beta)/w]^{\frac{1-\beta}{\beta}} LQ_f$, $Profits = \sum_{q_j \in \mathbf{q}_f} \pi q_j = \pi Q_f$, and $Production\ workers = \sum_{q_j \in \mathbf{q}_f} l_j = [(1-\beta)/w]^{\frac{1}{\beta}} LQ_f$.

making a major innovation conditional on having a successful innovation for a firm with n product lines. Then, $\mathcal{M}(n)$ can be expressed by

$$\mathcal{M}(n) \equiv \frac{x \sum_{k=0}^{\hat{k}} \theta (1-\theta)^k}{x+nz} = \frac{x\left[1-(1-\theta)^{\hat{k}+1}\right]}{x+nz}.$$

$\mathcal{M}(n)$ is a strictly decreasing function.

This result suggests that small firms and new entrants have a comparative advantage for achieving major advances. Large incumbents endogenously spend effort on maintaining and expanding existing products. Thus, while firms of all sizes obtain major advances, these major advances account for a smaller share of achieved innovations among larger firms. An important distributional implication of Proposition 5 is that these differences weaken when considering progressively larger thresholds $s_{\hat{k}}$. The comparative advantage is weakest at the most extreme values (i.e., $s_{\hat{k}=0} = \eta$).[12]

We empirically estimate these predictions in Section 5.1 and then use these results in our quantitative analysis. The baseline model makes many more predictions that we catalogue in Appendix B and investigate further in our NBER working paper.

3 Generalized Model

This section generalizes the innovation production function of the benchmark model. In particular, we assume that the production function for external innovations takes the form

$$X_n = \chi \left[R_x/\bar{q}\right]^{\psi} n^{\sigma}. \tag{33}$$

This production function nests two special forms. First, when $\sigma = 1-\psi$, the model becomes the extended Klette and Kortum (2004) framework where both internal and external investments scale up with firm size on a one-for-one basis with added product lines. Second, when $\sigma = 0$, we are back to the benchmark model of Section 2. We describe here the solution of the model under this generalized production function, and Section 6 quantifies this model and the σ parameter.

The static equilibrium of this generalized model follows exactly as the benchmark model, therefore we skip it (equations (11) – (15) hold identically). Moreover, when $\sigma > 0$, a firm that loses all of its product lines exits the economy. As we are not seeking analytical results, but instead preparing the general model for quantification, we eliminate the fixed cost and set $\Phi = 0$.

3.1 Research and Development by Incumbents

The production function in (33) delivers the R&D function

$$R_x = \bar{q}\tilde{\chi} n^{\tilde{\sigma}} x_n^{\tilde{\psi}},$$

[12] The aggregate quantity of major innovations by small and large firms depends upon these propensities and the firm size distribution.

where $x_n \equiv \frac{X_n}{n}$ is the innovation intensity per product line and

$$\tilde{\sigma} \equiv \frac{1-\sigma}{\psi}, \quad \tilde{\chi} \equiv \chi^{-\frac{1}{\psi}}, \quad \text{and} \quad \tilde{\psi} \equiv \frac{1}{\psi}.$$

In this case, the value function can be expressed as follows.

Proposition 6 *For a firm that has a quality portfolio* $\mathbf{q}$, *the value function has the following form:*

$$V(\mathbf{q}, \bar{q}) = A \sum_{q_j \in \mathbf{q}} q_j + B_n \bar{q}$$

where

$$(r+\tau) A = \pi + A^{\frac{\hat{\psi}}{\hat{\psi}-1}} \left[\frac{\lambda}{\hat{\psi}}\right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} \left(\hat{\psi}-1\right) \hat{\chi}^{\frac{1}{1-\hat{\psi}}}, \tag{34}$$

and

$$B_{n+1} = \left[\frac{(\rho + n\tau) B_n - n\tau B_{n-1}}{\tilde{\psi}-1}\right]^{\frac{\tilde{\psi}-1}{\tilde{\psi}}} \tilde{\psi}\tilde{\chi}^{\frac{1}{\tilde{\psi}}} n^{\frac{\tilde{\sigma}-\tilde{\psi}}{\tilde{\psi}}} + B_n - A[1+\bar{s}]. \tag{35}$$

Moreover, the optimal innovation efforts are defined as

$$z_j = \left[\frac{A\lambda}{\hat{\psi}\hat{\chi}}\right]^{\frac{1}{\hat{\psi}-1}} \text{ and } x_n = \left[\frac{A[1+\bar{s}] + B_{n+1} - B_n}{\tilde{\psi} n^{\tilde{\sigma}-1} \tilde{\chi}}\right]^{\frac{1}{\tilde{\psi}-1}}. \tag{36}$$

In this generalized model, the value function consists of two parts. The first part, which is denoted by A, is related to the discounted sum of future profits and internal innovations. By owning the product line, the firm will collect flow profits of πq_j until it is replaced at the rate τ. In addition, the firm can improve its quality q_j through internal innovations at the rate z_j, which also provides value to the firm. The second part, which is denoted by B_n, relates to the firm's external innovation capacity. By owning a product line, the firm has a franchise value of extending into new product lines through external innovations, which happens at the rate x_n. Since the production function is dependent on the number of product lines, this franchise value now is a function of n as well. The Klette and Kortum (2004) model assumes $B_n = nB$, while the baseline model of Section 2 assumes $B_n = B$.

Accordingly, the new flow equations for the fraction of firms with n product lines:

State :	Inflow	Outflow
$n = 0:$	$F\mu_1\tau =$	x_e
$n = 1:$	$F\mu_2 2\tau + x_e =$	$F\mu_1(2x_2 + \tau)$
$n \geq 2:$	$F\mu_{n+1}(n+1)\tau + F\mu_{n-1}(n-1)x_{n-1} =$	$F\mu_n(nx_n + n\tau)$

This summarizes the generalized model. Before we proceed a final remark is in order.

Remark 1 *Proposition 6 shows that innovation intensity* x_n *can be expressed as* $x_n = n^{\xi} f(n)$, *where*

$$\xi \equiv \frac{\psi + \sigma - 1}{1 - \psi} \tag{37}$$

and n^{ξ} *captures the direct effect on* n *on* x_n. *Note that* $f(n) = \left[\frac{A[1+\bar{s}]+B_{n+1}-B_n}{\tilde{\psi}\tilde{\chi}}\right]^{\frac{1}{\tilde{\psi}-1}}$ *captures the indirect effect of number of product lines on* x_n *through its impact on the franchise value* B_n. *When* $\psi + \sigma = 1$, *our model mirrors Klette and Kortum (2004) with* $f(n)$ *equal to some constant, whereas innovation intensity will be decreasing in firm size when* $\psi + \sigma < 1$. *Therefore* $\psi + \sigma$ *dictates the amount of decreasing innovation intensity in firm size.*

4 Patent Citation Behavior and Innovation Spillover Sizes

We now incorporate patent citation behavior across innovations into our benchmark model. As we have already defined the economy's equilibrium, our specified citation behavior does not affect real outcomes. We undertake this extension, however, to derive the economic meaning behind patent citations. This in turn allows us to quantify the model using richer data. Second, this addition demonstrates how this class of endogenous growth models captures many important features from empirical literature on patent counts and citations.[13] Trajtenberg (1990) is a well-known example of empirical work linking citations to economic value of innovations, albeit with noise. Constructing this link between these literatures is the central purpose of this section. Finally, this extension could provide a useful theoretical framework for future studies on the impact of policies such as intellectual property rights for innovation qualities and growth.

4.1 Forward Patent Citations

Innovations are clustered in terms of their technological relevances. Major innovations generate new technology clusters that last until they are overtaken by a subsequent major innovation. An example of the sequential innovation process was illustrated in Example 1 in Section 2.2.

Let $m(j,t)$ be the number of patents in the active technology cluster in product line j. For instance, if t is between the innovation times of P_3 and P_4 in the Example 1, then $m(j,t) = 3$, or if t is between P_{11} and P_{12}, then $m(j,t) = 2$. Therefore the number of citable patents in active technology clusters at time t is $M(t) = \int_0^1 m(j,t)\,dj$.

We next describe the citation distribution of patents. We specify citation behavior with a few simple rules that build upon the patent literature. Patents cite previous patents within the same technology cluster to specify how they build upon the prior work and the boundaries of the innovations. Each new patent, by definition, improves the previous technologically relevant innovations on some dimensions. However, not all subsequent innovations improve an existing technology in the same direction. Therefore major patents with broader scope are more likely to be cited by subsequent follow-on work (e.g., Lerner, 1994). We proxy this patent scope by the step

[13] Hall et al. (2001) provide a comprehensive introduction to patent citations. See also Hall et al. (2005), Jaffe et al. (2000), Jaffe et al. (1993), Thompson and Fox-Kean (2005), and Trajtenberg (1990).

size $s \in \{\lambda, \eta\alpha^k \mid k \in \mathbb{N}_0\}$ in our model. We assume that an innovation with size s will receive a citation from a subsequent patent within the same technology cluster with probability $s\gamma$ where $\gamma \in (0, 1/\eta)$. Finally a major innovation replaces the previous cluster. Thereafter, future citations begin with the new major innovation. Empirically, Hall et al. (2001) and Mehta et al. (2010) quantify the decline in relative citation rates over patent age that this model structure provides.

Thus, the citation behavior in Example 1 would be:

Cited	**probability**	**Citing**	**Cited**	**probability**	**Citing**
P_1 :	$\gamma\eta$	$P_2 - P_6$	P_6 :	$\gamma\eta\alpha^3$	*none*
P_2 :	$\gamma\eta\alpha$	$P_3 - P_6$	P_7 :	$\gamma\eta$	$P_8,\ P_9$
P_3 :	$\gamma\eta\alpha^2$	$P_4 - P_6$	P_8 :	$\gamma\lambda$	P_9
P_4 :	$\gamma\lambda$	$P_5,\ P_6$	P_9 :	$\gamma\eta\alpha$	*none*
P_5 :	$\gamma\lambda$	P_6	P_{10} :	$\gamma\eta$	P_{11}, P_{12} ...

4.2 Invariant Distributions

With these simple modeling assumptions, we can characterize the flow properties of citation behavior. These traits depend upon the real side of the economy and provide a richer description of it. Similar to our earlier expressions, the equilibrium of the economy requires an invariant citation distribution. Let $\Upsilon_{s_k,n}$ and $\Upsilon_{\lambda,n}$ denote the share of patents that are of size $\eta\alpha^k$ and λ, respectively, and receive n citations such that $\Sigma_{n=0}^{\infty}\Upsilon_{\lambda,n} + \Sigma_{k=0}^{\infty}\Sigma_{n=0}^{\infty}\Upsilon_{s_k,n} = 1$. For any given innovation size $s_k = \eta\alpha^k$, the flow equations for external patents with n citations take the following form

$$\begin{aligned} &\textsc{Outflow} && \textsc{Inflow} \\ M\Upsilon_{s_k,0}\tau\theta + M\Upsilon_{s_k,0}\gamma\eta\alpha^k\left(\tau\left(1-\theta\right)+z\right) &= \Psi_{k-1}\tau\left(1-\theta\right) \text{ for } n=0. && (38) \\ M\Upsilon_{s_k,n}\tau\theta + M\Upsilon_{s_k,n}\gamma\eta\alpha^k\left(\tau\left(1-\theta\right)+z\right) &= M\Upsilon_{s_k,n-1}\gamma\eta\alpha^k\left(\tau\left(1-\theta\right)+z\right) \text{ for } n \in \mathbb{Z}_{++}. && (39) \end{aligned}$$

The first line represents size s_k innovations with no citations ($n = 0$). There are $M\Upsilon_{s_k,0}$ such patents for each innovation size s_k. The first part of the outflow occurs when the technology cluster is replaced through a new major innovation at the rate $\tau\theta$. When this happens, patents become defunct and are no longer considered for citation. The second part of the outflow occurs when patents receive a new citation from subsequent innovations at the rate $\gamma\eta\alpha^k\left(\tau\left(1-\theta\right)+z\right)$. This latter expression is the probability of citation based on step size of $\gamma\eta\alpha^k$ multiplied by the arrival rate of subsequent patents. In this case, patents remain active but move up the citation distribution.

On the right hand side, the inflow occurs from Ψ_{k-1} product lines where the latest follow-up innovation was of size $\eta\alpha^{k-1}$ and a new follow-up innovation brings the product line into the Ψ_k group. This occurs at rate $\tau\left(1-\theta\right)$. This inflow is not dependent on the number of citable patents M. All patents initially have zero citations, and only a single patent can arrive per product line at any instant. The inflow thus depends only on the affected product lines.

Similar reasoning applies to the second row, where citations $n \geq 1$, except that the inflow occurs only from the $(k, n-1)$ group. These innovations arrive at rate $\tau(1-\theta)+z$ and they cite the specific patent at rate $\gamma\eta\alpha^k$.

Next we characterize the citation distribution of internal patents with flow equations

$$\begin{array}{rcll} & \textsc{Outflow} & & \textsc{Inflow} \\ M\Upsilon_{\lambda,0}\tau\theta + M\Upsilon_{\lambda,0}\gamma\lambda\left(\tau\left(1-\theta\right)+z\right) & = & z \text{ for } n=0 & (40) \\ M\Upsilon_{\lambda,n}\tau\theta + M\Upsilon_{\lambda,n}\gamma\lambda\left(\tau\left(1-\theta\right)+z\right) & = & M\Upsilon_{\lambda,n-1}\gamma\lambda\left(\tau\left(1-\theta\right)+z\right) \text{ for } n \in \mathbb{Z}_{++}. & (41) \end{array}$$

These flows have similar interpretation. The substantive difference is that the inflow of zero-cited patents occurs at rate z for internal improvements. The next proposition provides the explicit form solutions for these distributions.

Proposition 7 *The invariant distribution of the total number of forward citations (n) given to a patent of size $s \in \{\lambda, s_k \mid k \in \mathbb{N}_0\}$ can be expressed as*

$$\Upsilon_{s,n} = \Upsilon_{s,0}\Omega_s^n \text{ for } n \in \mathbb{N}_0,$$

where $M = \frac{x+z}{x\theta}, \Upsilon_{s_k,0} = \frac{\theta(1-\theta)^k\tau}{M[\tau\theta+\gamma s_k(\tau(1-\theta)+z)]}, \Upsilon_{\lambda,0} = \frac{z}{M[\tau\theta+\gamma\lambda(\tau(1-\theta)+z)]}$ *and* $\Omega_s \equiv \frac{\gamma s(\tau(1-\theta)+z)}{\tau\theta+\gamma s(\tau(1-\theta)+z)}$.

Similarly, the invariant distribution of the total number of external forward citations is

$$\tilde{\Upsilon}_{s,n} = \tilde{\Upsilon}_{s,0}\tilde{\Omega}_s^n \text{ for } n \in \mathbb{N}_0,$$

where $\tilde{\Upsilon}_{s_k,0} = \frac{\theta(1-\theta)^k\tau}{M[\tau\theta+\gamma s_k\tau(1-\theta)]}, \tilde{\Upsilon}_{\lambda,0} = \frac{z}{M[\tau\theta+\gamma\lambda\tau(1-\theta)]}$ *and* $\tilde{\Omega}_s \equiv \frac{\gamma s\tau(1-\theta)}{\tau\theta+\gamma s\tau(1-\theta)}$.

Note that $\Upsilon_{s,n}$ generates a more highly skewed distribution of citations as the share $\tau\theta$ gets smaller in the denominator. This is intuitive given the slower arrival of new technology clusters in favor of follow-on inventions that cite prior work.

5 Empirics of Innovation

This section provides empirical evidence regarding innovation and the firm size distribution that inform our model. We specifically focus here on regularities that will discipline the quantitative analysis in the next section, with further empirics to come when we can compare the quantified model and empirical data on untargeted dimensions.

5.1 Data Development

Our project employs the Longitudinal Business Database and the NBER Patent Database. The Longitudinal Business Database (LBD) provides the backbone for our research. This business registry contains annual observations for every private-sector establishment with payroll from 1976 onward (Jarmin and Miranda, 2002). The Census Bureau data are an unparalleled laboratory for studying the firm size distribution, entry/exit rates, and life cycles of U.S. firms. Sourced from U.S. tax records and Census Bureau surveys, the micro-records document the universe of establishments and firms rather than a stratified random sample or published aggregate tabulations. As a representative year, the data include 108 million workers and 5.8 million establishments in 1997.

We match into the LBD the individual records of all patents granted by the United States Patent and Trademark Office (USPTO) from January 1975 to May 2008. Each patent record provides information about the invention and the inventors submitting the application. Hall et al. (2001) provide extensive details about these data, and Griliches (1990) surveys the use of patents as economic indicators of technology advancement. We only employ patents 1) filed by inventors living in the United States at the time of the patent application, and 2) assigned to industrial firms. In 1997, this group comprised about 77 thousand patents (40% of the total USPTO patent count in 1997, with most of the residual being patents to foreign inventors). We match these patent data to the LBD using firm name and location matching algorithms that build upon Balasubramanian and Sivadasan (2011) and Kerr and Fu (2008).[14]

Our final sample retains firms that are ever granted a patent by the USPTO, and we have 23,927 firms over the 1982-1997 period, using earlier periods for metrics like backwards self-citation shares and later periods for forward growth rates and similar. There are two very important features about this dataset to highlight. First, our sample only includes innovative firms, which have a different firm size distribution than the economy as a whole. In our sample, for example, 14% of firms have more than 500 employees at some point in their life span (12% for all observations of the firm), while this share is less than 1% for the whole economy. This tilt towards larger firms is not surprising, and as discussed in detail by Hurst and Pugsley (2011), the majority of small firms do not target innovation or growth. We thus exclude non-innovative firms from our sample to be in keeping with the model of innovative firms and the heterogeneity in their types of innovations, and our firm moments should be viewed in this context. Nevertheless, we are building upon the universe of patenting firms and include many small firms in the sample.[15]

Second, only a few innovative firms patent in every year, and the same is true in our model with respect to realizing an innovation. These considerations lead us to use our data in two ways. In some cases (e.g., Gibrat's law estimations), we conduct an annual analysis as the necessary ingredients are continually observed in both the data and the model. In other cases (e.g., quality distributions of realized innovations), we focus on five-year periods and the firms achieving innovations as depicted

[14] Our NBER working paper describes this matching procedure and the data employed more extensively. The working paper also provides complementary evidence from the National Science Foundation's R&D Survey that supports the patent-based results provided here. The NSF Survey sub-samples R&D performers that conduct less than $1 million in R&D annually, and thus our focus on patenting allows us greater confidence for capturing the complete firm size distribution for innovative firms.

[15] Approximate 25th, 50th, and 75th percentile levels are 17, 70, and 370 employees. These are "fuzzy" averages around these points in order to satisfy Census Bureau disclosure requirements. The mean employment level is about 1805 workers.

below. It is important to emphasize in advance that our quantitative model exactly mirrors each data development step described below to ensure that we precisely align the model with the data. This mirroring technique has the powerful advantage of allowing us to select the approach that best suits each prediction, accounting for the nuance of the data assembled.

5.2 Firm Growth by Firm Size

Our first model prediction is that small firms grow faster than large firms.[16] We test this prediction using annual employment growth patterns in our sample of U.S. innovative firms. Following Lentz and Mortensen (2008), we define for firm f the employment growth of $EmpGr_{f,t} = [Emp_{f,t+1} - Emp_{f,t}]/Emp_{f,t}$. We model employment growth without conditioning on survival and thus retain $EmpGr_{f,t} = -1$ for businesses that close between t and $t+1$ (the LBD measures employment in March of each year). This metric is unbounded upwardly, and we impose a 1000% growth cap. With this winsorization, the mean of $EmpGr_{f,t}$ is 0.0745. Controlling for industry-year fixed effects $\eta_{i,t}$, we estimate[17],

$$EmpGr_{f,t} = \eta_{i,t} - \underset{\text{(s.e. 0.0013)}}{0.0351} \cdot \ln(\text{Emp}_{f,t}) + \epsilon_{f,t}.$$

This coefficient finds a 10% increase in firm employment is associated with a 0.35% reduction in forward employment growth, or about 5% of the sample mean. The growth impact of the interquartile range of firm size (approximately 17 to 370 employees) is 10.8%, somewhat larger than the mean. This relationship is very robust to alternative measures of firm size (e.g., establishment counts, industry groups), weighting observations, or considering panel variation, reflecting the many settings where it has been observed in prior research. Conditional estimations that exclude exiting firms yield a steeper negative relationship, as does raising the maximum growth rate (discussed further below in model robustness checks). When using the Davis et al. (1996) formula that compares growth to the average of the two periods, the conditional estimation yields a negative relationship, while the unconditional estimation is inconclusive across variants.

5.3 Innovation Intensity by Firm Size

We next study the innovation intensity to firm size relationship. Our model expresses this relationship in two ways. It first appears in research input terms via the R&D-to-sales investments made by firms; later, the relationship is realized as differing innovation output intensities that follow-on from

[16]The empirical deviation from Gibrat's Law of proportionate growth is extensively documented in surveys such as Sutton (1997), Caves (1998), and Geroski (1998) and is among the stylized facts in Klette and Kortum (2004). The Klette and Kortum (2004) model yields Gibrat's law. Lentz and Mortensen (2008) show that the addition of firm heterogeneity into the Klette and Kortum (2004) model is consistent with deviations from proportionate growth observed in Danish firm-level data.

[17]The regression sample includes 146,678 observations. We assign industries to firms at the two-digit level of the Standard Industrial Classification system using industries in which firms employ the most workers. Regressions are unweighted and cluster standard errors at the firm level.

these investments and firm capabilities. We can discipline the model through either relationship, and for several data quality reasons we pursue the realized rate of innovation outputs.[18]

We study this prediction through patents per employment $Patent/Empl_{f,t}$, where the timing of patents is by their application year. The largest innovative firms like Microsoft or Boeing apply for many patents each year, but most innovative firms are irregular and lumpy in their patent filings. We thus analyze this prediction with five-year periods that extend 1982-1986, 1987-1992, and 1992-1996. (With some abuse of notation, we continue to use t to represent time periods.) We focus this exercise on "continually innovative firms" in the sense that included firms file at least one patent in each five-year period that they are observed to be in operation. This dataset includes 16,818 firm-period observations. The continuous sample approach keeps a consistent definition with respect to non-zeros and facilitates a sharper match with the model, where we also impose this requirement for included firms to be continually innovative over five-year periods.

To prepare for the future matching of our data moment to the model, we transform $Patent/Empl_{f,t}$ to be of mean zero and unit standard deviation during each period. We use the transformed series because the exact level of U.S. patenting per employee does not have a direct meaning or counterpart to the model's levels. By placing both data and model outcomes into unit standard deviations, we are able to match and compare them.

Our key estimation is

$$Patent/Empl_{f,t} = \eta_{i,t} - \underset{\text{(s.e. 0.0058)}}{0.1816} \cdot \ln(\text{Emp}_{f,t}) + \epsilon_{f,t}.$$

This coefficient finds a 10% increase in firm employment is associated with a reduction of 0.018 standard deviations in patents per employee among innovative firms. Across the interquartile range of firm sizes, the impact is 0.561 standard deviations. If we relax the continuous innovator sample restriction, the coefficient is very similar at -0.164. We also find robust results with the many regression variants discussed above with the employment growth specifications.

5.4 Fraction of Major Innovations by Firm Size

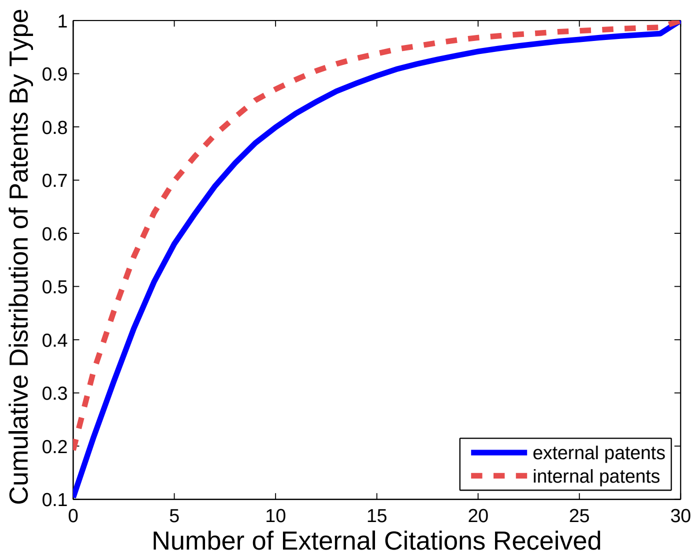
If external innovations have a higher average step size than internal innovations, then our baseline model makes a very important prediction that small innovative firms and new entrants have a comparative advantage for achieving major advances. In our framework, large incumbents spend proportionately more effort on maintaining and expanding existing products that they own. The

[18] Our NBER working paper tabulates R&D expenditures per sales or per employee across the firm size distribution using the NSF R&D Survey. While these tabulations accord with the theory as applied to innovative firms, the R&D data are limited by their subsampling of small R&D performers. Sales information for U.S. firms is also limited to years in which the economic census is conducted. We thus provide evidence of this prediction instead through patents per employment as we observe the universe of both of these data elements, providing confidence for the comprehensive nature of our estimation. Regardless of how measured, these innovation intensity to firm size relationships depend in clear ways on our isolation of innovative firms compared to the whole economy, as we are excluding the many small businesses that are not engaged in innovative work. This is in keeping with our model of innovative firms and the evolving quality ladders of innovations.

model's solution does not require internal or external innovations to be larger in average step size, but this prediction emerges if the larger step size for external innovation holds.

Figure 1 provides some necessary empirical evidence regarding the relative step sizes of external vs. internal work using patent citations. The sample is restricted to industrial firms that have all inventors located in the United States. We plot the distribution of external citations (i.e., excluding self citations) received for patents filed between 1975-1984. The solid line represents patents that make no backwards citations to prior work of the assignee at the time of filing, a proxy for external innovations. The dashed line represents patents whose internal backwards citations are the majority of their citations, a proxy for internal R&D. Both series display a large number of patents with no external citations and a skewed distribution, which are predictions of our framework. More important, the comparison of the external and internal distributions shows that the former exceeds the latter in a form akin to first-order stochastic dominance.[19]

Figure 1: Citation Distribution by Patent Type



With this background, we next verify that small innovative firms and new entrants have a comparative advantage for achieving major advances. We first identify the quality of each patent in terms of its external citations compared to its peers from the same technology class and application year. Constructing an indicator variable for the patent being in top decile in terms of these external citations, we calculate $TopPatentShare_{f,t}$ as the average of these patent-level indicators across a time period for a firm. Not surprisingly, the average of this variable is about 0.10. We then estimate

[19]The differences are statistically significant and hold in regressions that control for a variety of traits about the patents (e.g., technology-year fixed effects) or firm fixed effects. The omitted, middle group (i.e., patents where backwards self citations are present but not a majority) behaves similarly to the no self citation group and are excluded for visual clarity; later, we will group them with external patents for our model quantification.

this firm-level measure as a function of the size distribution as

$$TopPatentShare_{f,t} = \eta_{i,t} - \underset{\text{(s.e. 0.0008)}}{0.0034} \cdot \ln(\text{Emp}_{f,t}) + \epsilon_{f,t}. \tag{42}$$

This estimation finds that a 10% increase in firm employment is associated with a reduction of 0.034% in the fraction of a firm's patents among the top decile of the patent quality distribution. Relative to the sample mean, this effect is 0.34%. Across the interquartile range of firm sizes, the impact is 0.011, or a tenth of the sample mean.

Table 1 broadens the lens and repeats specification (42) for each quartile of the patent quality distribution using our continuous innovation sample. The first column documents the lowest quality quartile, while the last column is the highest one; coefficients across the four specifications naturally sum to zero. Estimations again control for industry-period fixed effects. Larger firms are associated with a systematic shift in the quality of their patents out of the top quartile and into the bottom half of the distribution. Our working paper further uses this framework to confirm our hypothesis that firm size differentials weaken with more-stringent citation quality thresholds due to the increasing relative importance of the stochastic nature of realized inventions.

TABLE 1: FIRM SIZE AND PATENT QUALITY DISTRIBUTION

	Share of firm's patents in quality distribution range:			
	[0,25)	[25,50)	[50,75)	[75,100]
Log firm employment$_t$	0.0027	0.0048	0.0000	-0.0074
	(0.0009)	(0.0010)	(0.0010)	(0.0012)

Notes: Estimates include 16,818 observations, are unweighted, and cluster standard errors by firm.

6 Quantitative Analysis

We estimate our model using micro data described in Section 5.1. Section 6.1 outlines the computational solution of the generalized model. Section 6.2 describes our identification strategy. Section 6.3 provides the main estimation results, and Section 6.4 provides robustness checks.

6.1 Computer Algorithm

We solve the generalized model as a fixed point over the growth rate g. Our algorithm employs a computational loop with the following steps:

1. Guess a growth rate g.
   (a) Guess a creative destruction rate τ.
      i. Solve for A in (34), the sequence $\{B_n\}$ in (35), and z_j and $\{x_n\}$ in (36).
      ii. Verify the free entry condition as a function of $\tau : A\,[1 + \bar{s}] + B_1 = \nu$.
      iii. If not converged, update τ and go to step 1(a)i.

   (b) Calculate the growth rate: $g = \tau \bar{s} + z\lambda$.

   (c) Update the growth rate. If not converged, go to step 1a.

2. End the equilibrium solver.

3. Simulate a sample of firms and compute the moments of interest.

The sequence of firm value functions in step 1(a)i is solved using the uniformization method (see Acemoglu and Akcigit, 2012 for details). In step 3, we simulate a sample of 2^{14} firms (16,384) and iterate for 500 years until we obtain convergence. At each iteration, firms gain and lose products according to the flow probabilities specified in the model.

6.2 Identification

Our model has 13 structural parameters: $\sigma, \rho, \theta, \eta, \alpha, \lambda, \beta, \nu, \gamma, \hat{\psi}, \tilde{\psi}, \hat{\chi}, \tilde{\chi}$. We identify these parameters in three ways. First, we fix three parameters $(\rho, \hat{\psi}, \tilde{\psi})$ using values developed in Section 6.2.1 from the literature and R&D-based regressions. Second, we use the observed distribution of citations for patents to pin down three elements of the step size distribution $(\theta, \alpha, \eta\gamma)$ in Section 6.2.2. Finally, for the remaining parameters and to parse $\eta\gamma$, we target the relevant firm moments in the data. One critical part of this third step is to identify the key decreasing returns parameter σ using an indirect inference approach, where we replicate the regressions of Sections 5.2-5.3 using data simulated from the model.

6.2.1 Externally Calibrated Parameters

We set the discount rate equal to $\rho = 2\%$, which roughly corresponds to an annual discount factor of 97%.

We rely on prior literature for estimates of the curvature of the R&D cost function, which we will set equal across internal and external innovation $\hat{\psi} = \tilde{\psi}$ (the model retains shifters in these cost functions). One line of studies quantifies the elasticity of patents to R&D expenditures (e.g., Griliches, 1990, Blundell et al., 2002, Hall and Ziedonis, 2001). This literature often concludes this elasticity is around 0.5, which implies a quadratic curvature $\psi = 2$. Acemoglu et al. (2013) reach a similar estimate using the Census Bureau data as well when focusing on firms in the R&D Survey. The second set of papers examines the impact of R&D tax credits on the R&D expenditure of firms (e.g., Hall, 1992, Bloom et al., 2002, Wilson, 2009. In a survey of this work, Hall and Van Reenen (2000) conclude that a tax price elasticity of around unity is typically found, which again corresponds to a quadratic cost function.[20] Given this common finding, we set $\hat{\psi} = \tilde{\psi} = 2$.

[20] The mapping to our setting is straightforward. To simplify the notation, let us denote a single R&D spending relationship $R = P x_n^{\psi} F_n$, where P is the price of R&D and F_n is a multiplicative term that can potentially depend on firm size. If the return to innovation is Π, the generic maximization problem can be written as $\max_{x_n} \left\{ x_n \Pi - P x_n^{\psi} F_n \right\}$. Solving for the first order condition, $R = P^{-\frac{1}{\psi-1}} F_n^{-\frac{1}{\psi-1}} [\Pi/\psi]^{\frac{\psi}{\psi-1}}$. Hence the price elasticity of R&D spending in our model corresponds to $d \ln R / d \ln P = -\frac{1}{\psi-1}$. A unitary estimate corresponds to $\psi = 2$.

Section 6.4.4 will study the robustness of the results with alternative R&D elasticities of 0.4 and 0.6.

6.2.2 Citation Distribution

Our model yields an analytical solution for the patent citation distribution that is dictated by the innovation step-size parameters. In particular, when we focus only on external citations ($z_j = 0$), the distribution of patents that are of quality s_k and receive n citations is simply

$$\Upsilon_{s_k,n} = \Upsilon_{s_k,0}\Omega_{s_k}^n \text{ for } n \in \mathbb{N}_0,$$

where $\Upsilon_{s_k,0} = \frac{\theta^2(1-\theta)^k}{\theta+\gamma\eta\alpha^k(1-\theta)}$ and $\Omega_{s_k} \equiv \frac{\gamma s_k(1-\theta)}{\theta+\gamma s_k(1-\theta)}$. $\Upsilon_{s_k,n}$ gives us the joint distribution of patents that are k-times incremented and have received n citations. Our model provides the analytical distribution of k-times incremented patents from (18) as $\Psi_k = \theta(1-\theta)^k$ for $k \geq 0$. Hence, we can find the marginal distribution of n-times cited patents as

$$\mathcal{F}_n(\theta, \gamma, \eta, \alpha) = \sum_{k=0}^{\infty} \Psi_k \Upsilon_{s_k,n}.$$

The empirical tractability comes from the fact that the distribution of n-times cited patents depends only on four structural parameters: $\theta, \gamma, \eta, \alpha$. Citation distributions do not allow one to distinguish between the overall quality level of external inventions (η) and factors that govern the general tendency of patents to cite each other (γ). Since γ and η always appear multiplicatively in the shape of the citation distribution, we can use these data to identify the three parameters of θ, α, and the combination of $\eta\gamma$.

Figure 2: Citation Distribution

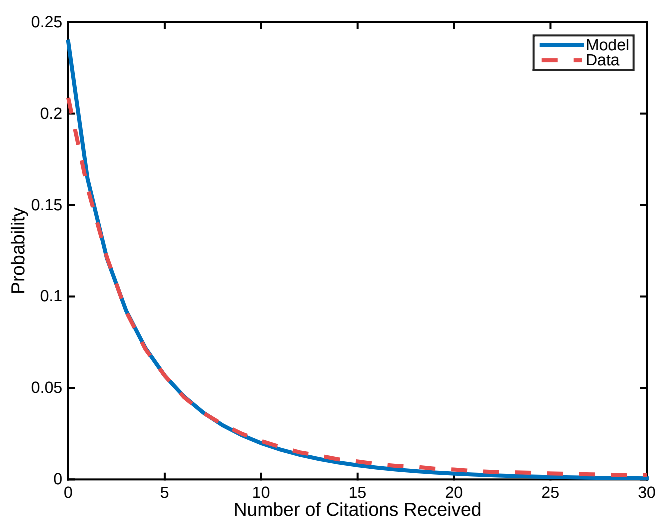


Figure 2 plots the empirical distribution together with the model-generated citation distribution.

The model does a very good job in replicating the data.

Table 2 lists the resulting parameter estimates. Roughly 10% of external innovations are found to be significant enough to open new technology clusters, and the decay rate α for the quality of external work is fairly modest. The $\gamma\eta$ estimate suggests that patents that open a new technology cluster have a 75% probability of being cited by later patents in the cluster.

Table 2: Citation Distribution Parameters

θ	$\gamma\eta$	α
0.103	0.750	0.929

6.2.3 Indirect Inference

There are seven remaining parameters to be estimated: $\sigma, \tilde{\chi}, \hat{\chi}, \eta, \lambda, \beta, \nu$, which will also identify γ through Table 2. We identify these parameters using an indirect inference approach in the spirit of Lentz and Mortensen (2008). We compute various model-implied moments from the simulation strategy described above and compare them to the data-generated moments to minimize

$$\min \sum_{i=1}^{7} \frac{|\text{model}(i) - \text{data}(i)|}{\frac{1}{2}|\text{model}(i)| + \frac{1}{2}|\text{data}(i)|},$$

where we index each moment by i. Our indirect inference procedure targets seven moments that we describe next. The generalized model does not yield an analytical solution, and thus we cannot express the targeted moments in this form. However, we build intuition by using the analytical solutions to Section 2's benchmark model to guide us in choosing the right moments for identification. For ease of these depictions, we abstract from quality levels by setting $q_j = 1$, $\forall j$, although innovation qualities are clearly included in the simulation of the general model.

Average Profitability For both the benchmark and generalized models, the profit-to-sales ratio is equal to $\mathbb{E}\left(profit_f / sales_f\right) = (1-\beta)^{\frac{2\beta-1}{\beta}} \tilde{\beta}^{\frac{1}{\beta}}$, where $\tilde{\beta} \equiv \beta^{\beta}[1-\beta]^{1-2\beta}$. We therefore target the average profitability in the economy to help identify β. The profit-to-sales ratio in the model includes R&D expenditures, and thus we combine annual published BEA pre-tax profit rates with industrial R&D expenditure rates to determine an estimate of 10.9% for the 1982-1997 period.

R&D Intensity and Internal-to-External Citations Ratio We discipline the R&D scale parameters $\hat{\chi}$ and $\tilde{\chi}$ through measures of R&D intensity and the citation ratio of internal vs. external innovations. Aggregating across firms and using Proposition 4, the baseline model shows the economy-wide R&D-sales ratio to be a linear combination of $\hat{\chi}$ and $\tilde{\chi}$. This ratio is 4.1% in our sample.[21] In addition, the citation ratio of internal vs. external innovations informs the R&D scale

[21] For this purpose, we need to make use of the R&D Survey, which samples with certainty firms that conduct more than $1 million dollars of R&D and subsamples firms beneath this threshold. Our first step builds a sample of

parameters as

$$\frac{\lambda\tilde{\chi}}{(1+\bar{s})\,\hat{\chi}}.$$

We define internal patents as those with 50% or more of backwards citations being to assignees of the same firm. This approach is similar to Figure 1, with the explicit ten-year window from application date ensuring that the procedure is consistent across the sample period. We estimate this ratio using external citations to be 0.774 (= 5.023/6.488). These data inputs will inform the R&D scale parameters.

Fraction of Internal Patents and Aggregate Growth Rate Our model has four parameters that govern the step-size dynamics: $\theta, \alpha, \eta, \lambda$. We previously identified θ and α through the citation distribution. The remaining two parameters are the step size for internal innovations λ and the step size of radical innovations η. Step sizes determine both the innovation incentives and the aggregate growth rate:

$$z_j = \left[\frac{\lambda\nu}{(1+\bar{s})\,\hat{\psi}\hat{\chi}}\right]^{\frac{1}{\hat{\psi}-1}} \text{ and } g = \tau^*\bar{s} + z^*\lambda.$$

We can therefore discipline η and λ by targeting the fraction of internal patents $\left(\frac{z}{z+\tau}\right)$ and the growth rate. The internal patent share is 21.5%. The aggregate growth rate is calculated in deflated terms and on a per employee basis to match the model and the BEA profit estimates. This ranges from 0.91%-1.03% depending upon details of the calculation, and we assign a value of 1.0%.

Entry Rate The entry rate in the benchmark model is $x_e = \tau \exp(-x/\tau)$. Equations (22) and (24) show that the creative destruction rate is decreasing in the entry cost parameter ν, $d\tau/d\nu < 0$, and equation (23) shows incumbent efforts are increasing in entrant costs, $dx/d\nu > 0$. Therefore the impact of entry cost on the flow of entry is strictly negative, $dx_e/d\nu < 0$, and thus targeting the entry rate can help inform the entry parameter. The entry rate in our data is 5.82%, measured over five-year intervals through employments among patenting entrants.

Firm Growth vs Firm Size Regression from Section 5.2 The extended Klette and Kortum (2004) approach, where $\sigma = 1-\psi$, predicts that the unconditional firm growth would be independent of firm size, whereas the benchmark model with $\sigma = 0$ goes to the other extreme and predicts that firm growth is decreasing in firm size. In order to identify the actual value of σ, we mirror the same

firm-period observations for which we observe reported R&D, sales, and employment. The five-year periods match those of our core sample. We then merge in patents, including zero-valued outcomes. From this, we obtain an average conversion factor for relating R&D/sales to patents/employee. The second step applies this conversion factor to our full sample, where our aggregate patent/employee statistic includes firms that did not patent. This procedure gives us an aggregated value that closely aligns with other estimates of R&D/sales ratios. What is very important to emphasize for our subsequent work is that these values are determined through aggregates over the whole sample, not firm-level imputations. As the largest companies account for the substantial majority of these variables and will be surveyed directly by the R&D Survey, the procedures used here are quite robust.

growth-size regressions with data generated from the simulated model. The empirical coefficient of interest from the earlier work is -0.035.

6.3 Benchmark Estimation Results

Table 3 reports the empirical and simulated moments using the generalized model.

Table 3: Moments

Moment	Data	Model	Moment	Data	Model
profitability	0.109	0.106	entry rate	0.058	0.066
R&D intensity	0.041	0.042	average growth rate	0.010	0.010
internal/external cite	0.774	0.732	growth vs size (fact 1)	-0.035	-0.035
fraction of internal patents	0.215	0.250			

Overall, the model matches closely the targeted moments. In particular, when we replicate the same regression as in Section 5.2, we get the identical regression coefficient. The resulting parameter estimates are reported in Table 4.

Table 4: Estimated Model Parameters

σ	$\tilde{\chi}$	$\hat{\chi}$	η	λ	β	ν
0.395	4.066	0.346	0.112	0.051	0.106	0.830
Implied $\sigma + \psi = 0.895$.						

Our estimates find that there are some decreasing returns in firm size for external innovation as captured by the value of $\sigma \approx 0.4$. Among the other results, the ratio of $\tilde{\chi}$ to $\hat{\chi}$ suggests that the R&D cost parameter for external innovations is about 12-fold larger than for internal R&D. External innovations that open up a new technology cluster are estimated to have more than twice the potency of internal work. With the decay rate of $\alpha = 0.929$, roughly ten follow-on external innovations occur before this work is less valuable than internal innovations.

6.3.1 Characterization of the Economy

To provide further intuition on how σ plays a role in generating size-dependent firm moments, Figure 3 plots the franchise value function of a firm B_n as a function of the number of product lines n when $\sigma \in \{0, 0.2, 0.4, 0.5\}$. Figure 4 similarly plots the resulting external innovation intensity X_n. The franchise value function B_n for the baseline model in Figure 3 is flat because external innovation does not scale, while it grows linearly in the Klette and Kortum (2004) scenario. The small dashed line shows that the franchise value with $\sigma = 0.4$ grows similarly to the Klette and Kortum (2004) framework among smaller firms, with more modest departures after that. Figure 4 likewise illustrates that external innovation intensity declines with firm size but stabilizes in a way that limits the full dilution in the baseline model.

Figure 3: Franchise Value B_n

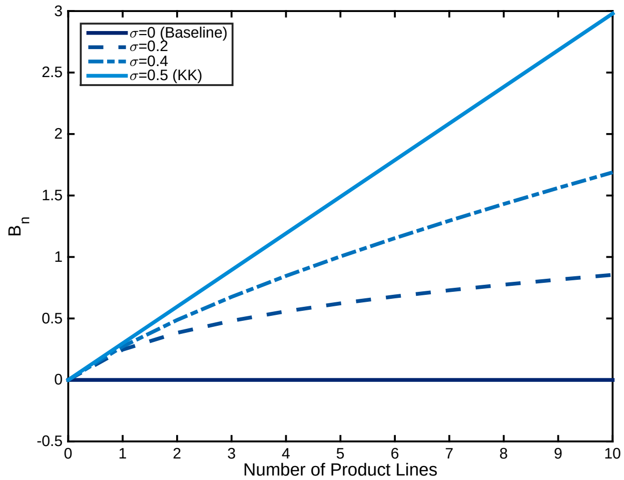
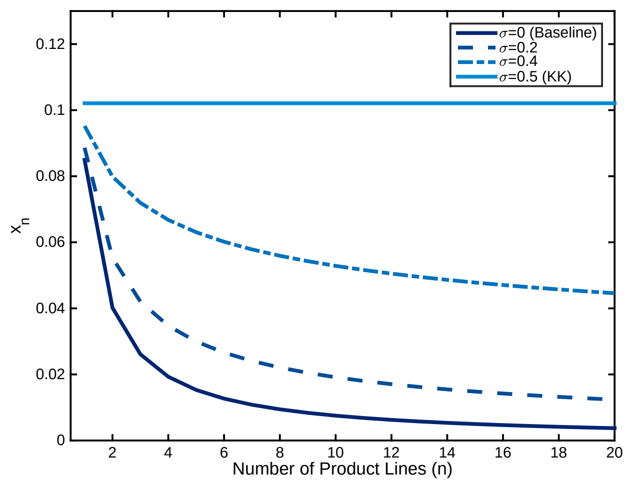


Figure 4: Innovation Intensity x_n



In our model, firm size is determined by the combination of the number of product lines and their quality distributions. Figure 5 illustrates the very tight correspondence of product lines to firm size in our model, with the latter normalized to average quality level in the economy, which builds additional connections and intuitions to the frameworks of Klette and Kortum (2004) and Lentz and Mortensen (2008).

Figure 5: Firm Size vs Number of Product lines

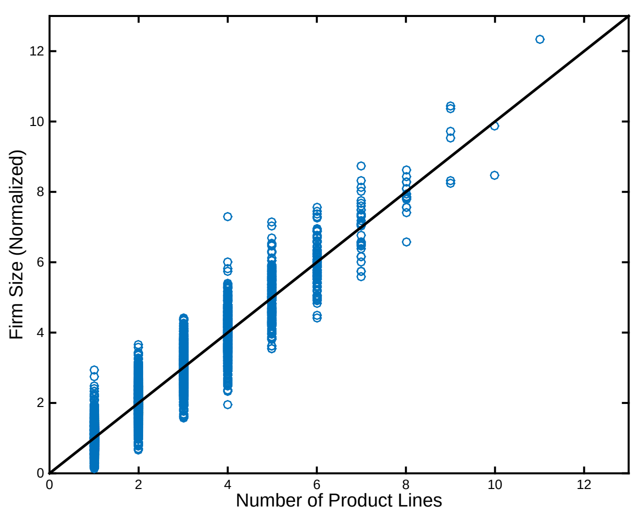


Figure 6 demonstrates that our framework generates an invariant product-line distribution at the firm level that resembles an exponential distribution. Combined with the quality margin, the invariant firm size distribution is illustrated in Figure 7.

Figure 6: Product Line Distribution

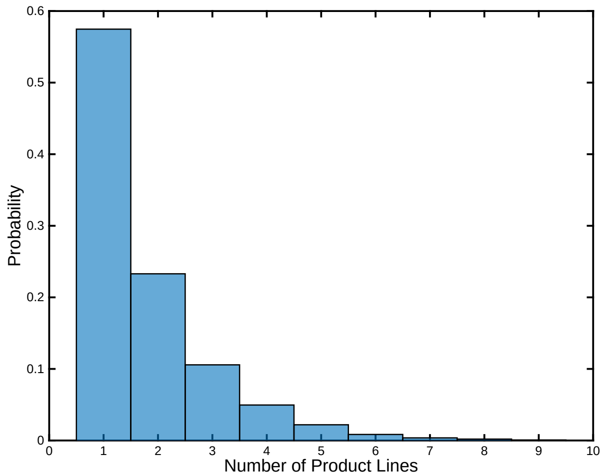
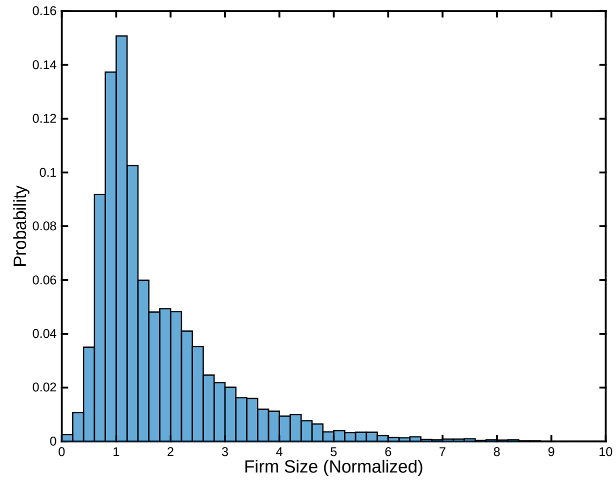


Figure 7: Firm Size Distribution



6.3.2 Growth Decomposition

We now use the structure of our model to document the sources of growth. In our model, growth is driven by (*i*) new entrants, (*ii*) incumbents doing internal innovations on their existing lines, and (*iii*) incumbents expanding into other lines through external innovations:

$$g = \underbrace{x_e \bar{s}}_{\text{entry}} + \underbrace{\sum_{n=0}^{\infty} F \mu_n X_n \bar{s}}_{\text{incumbent external}} + \underbrace{z\lambda}_{\text{incumbent internal}} .$$

Table 5 reports the magnitudes of each of these components in our model.

Table 5: Growth Decomposition

Actual Values			*In Percentage Terms*		
Internal	**External**	**New Entry**	**Internal**	**External**	**New Entry**
0.0020	0.0055	0.0026	19.8%	54.5%	25.7%

Our model estimates that 26% of aggregate productivity growth is driven by new entry. Of the three-quarters of productivity growth that comes from the action of incumbent firms, the majority of it depends upon external innovation efforts of firms. These figures are consistent with the empirical work that is surveyed by Foster et al. (2000), recognizing that some of our external innovation effect would be viewed as entry/exit in prior empirical calculations.

Another important distinction between external innovation and internal innovation is the differential impacts on qualities. The average step size associated with external innovations is $\bar{s} = 0.069$, whereas the step size of internal innovation is $\lambda = 0.051$, which implies that an average external innovation has 35% $(= 0.069/0.051 - 1)$ higher impact than internal innovation.

6.3.3 Comparison of Untargeted Moments

We next compare our quantified model against untargeted features of the data. We do this through nonparametric regressions that compare variables across the firm size distribution. We include indicator variables by firm size quintile, with the smallest firm size category serving as the reference group. Our model estimation only targets the annual linear relationship for firm size and growth, and so the degree to which we observe comparable patterns for other variables across the firm size distribution provides confidence in the model's performance. For the exercises, we use the continuous innovation sample in both datasets so that all variables are defined and the samples remain consistent over tests. We structure our model simulation such that the model-developed data ($n = 16,371$) has comparable statistical properties to our Census Bureau data ($n = 16,818$).[22]

Table 6 considers four main variables for which we have provided initial empirical evidence thus far.

Table 6: Firm size distribution and data-model comparison

	Growth rate to next period	Normalized patent per employee	Internal patent share	Top 10% patent share
	Panel A. Model, effects relative to smallest size quintile			
2nd quintile	-0.1284 (0.0210)	-0.8194 (0.0392)	-0.0134 (0.0114)	-0.0032 (0.0063)
3rd quintile	-0.2159 (0.0199)	-1.1065 (0.0379)	-0.0116 (0.0111)	-0.0055 (0.0060)
4th quintile	-0.3202 (0.0191)	-1.3404 (0.0372)	0.0256 (0.0105)	-0.0059 (0.0056)
Largest quintile	-0.3866 (0.0188)	-1.5507 (0.0368)	0.0538 (0.0099)	-0.0065 (0.0053)
	Panel B. Data, effects relative to smallest size quintile			
2nd quintile	-0.0133 (0.0502)	-0.9067 (0.0336)	0.0190 (0.0044)	-0.0030 (0.0078)
3rd quintile	-0.2790 (0.0464)	-1.0780 (0.0320)	0.0356 (0.0048)	-0.0211 (0.0076)
4th quintile	-0.2865 (0.0462)	-1.1166 (0.0322)	0.0413 (0.0047)	-0.0296 (0.0072)
Largest quintile	-0.4052 (0.0448)	-1.1351 (0.0323)	0.0471 (0.0045)	-0.0211 (0.0072)

Notes: Estimates are unweighted and cluster standard errors by firm.

On all four dimensions, the model closely matches the data in terms of the direction of differences across the firm size distribution: slower growth, lower patents per employee, higher share of patents being internal, and a lower share of patents being in the top 10% in terms of external impact. The model predicts a larger five-year growth differential between the smallest quintile and the second quintile than present in the data, but the differences for larger quintiles are quite similar. Patents per employee are very similar in levels and direction. The model under-predicts the initial rise in internal patent shares present in the data, but the effects for the largest quintiles are very close. Finally, the model under predicts the steepness of the decline in top/radical patents, but otherwise

[22] We continue to organize our sample around five-year blocks. The three periods included in the regressions are 1978-1982, 1983-1987, and 1988-1992, and we use earlier and later data to calculate variables as required. In estimations with Census Bureau data, we include $\eta_{i,t}$ fixed effects for the industry i and year t of the firm. Industries are assigned to firms at the two-digit level of the Standard Industrial Classification system using industries in which firms employ the most workers. All estimations cluster standard errors at the firm level and are unweighted.

shows a very similar coefficient pattern.[23] Overall, these results are very encouraging given that the model has not been targeting these firm size distribution components or time dimension.

Table 7 continues with this approach and considers the patent quality distribution more broadly. We calculate the share of patents for each firm-period that fall within the indicated quartile of the quality distribution. In the data, these quality distributions are measured through external citations relative to the application year and technology of the patent. The model again performs quite well in this untargeted test. Perhaps most striking, the model correctly predicts the disproportionate mass of patents for the largest firms falling within the second quality quartile, and it gets the relative size of this effect very close to the data. This part of the distribution is where internal patents sit and is a very distinctive piece of the framework developed in this paper. The model also correctly predicts that most of this extra mass is being shifted from the top quartile of external impact.[24]

Table 7: Firm size distribution and patent quality distribution comparison

	Share of firm patents in quality distribution range:			
	[0,25)	[25,50)	[50,75)	[75,100]
	Panel A. Model, effects relative to smallest size quintile			
2nd quintile	0.0039 (0.0091)	-0.0081 (0.0122)	-0.0133 (0.0094)	-0.0091 (0.0097)
3rd quintile	0.0111 (0.0090)	-0.0055 (0.0119)	-0.0051 (0.0090)	-0.0107 (0.0094)
4th quintile	0.0012 (0.0082)	0.0153 (0.0112)	-0.0028 (0.0083)	-0.0137 (0.0088)
Largest quintile	-0.0108 (0.0077)	0.0386 (0.0104)	-0.0045 (0.0078)	-0.0232 (0.0082)
	Panel B. Data, effects relative to smallest size quintile			
2nd quintile	-0.0079 (0.0079)	0.0054 (0.0090)	0.0074 (0.0095)	-0.0049 (0.0106)
3rd quintile	0.0039 (0.0081)	0.0317 (0.0093)	-0.0008 (0.0094)	-0.0349 (0.0105)
4th quintile	0.0122 (0.0078)	0.0405 (0.0090)	0.0025 (0.0092)	-0.0552 (0.0102)
Largest quintile	0.0140 (0.0074)	0.0327 (0.0080)	0.0037 (0.0086)	-0.0503 (0.0099)

Notes: See Table 6.

Table 8 finally compares firm-level growth regressions in the model and data. These tests evaluate whether the micro-dynamics of firms behave similarly as we consider all elements together. We use the continuous innovator samples and five-year periods. The central regressors to explain employment growth to the next period are the firm's current employment, the firm's total patenting in the period, the quality distribution of the firm's own patents in this period (Patent Quality Share$_{f,q}$), and the share of a firm's patents that are internal in nature (Internal Share$_{f,q}$).

[23] The model coefficients are not statistically different from zero for the last column. In unreported estimations, we develop a larger model sample of 152,089 data points, where we find a largest quintile impact of -0.0071 (0.0017). Thus, our attention focuses mainly on the coefficient magnitudes between the model and data, versus statistical precision. The complete results for Tables 6-8 with the larger sample are available upon request and are very similar to those reported.

[24] The largest firms in Panel B also show some modest mass at the lowest quartile. In the model, the constant internal step size λ concentrates the internal effect into a single quartile. The fact that we overall match the quality distribution so well indicates that this simplifying structure is a reasonable approximation.

Specifications take the form

$$EmpGr_{f,t} = \eta_{i,t} + \gamma_E \ln(\text{Emp}_{f,t}) + \gamma_P \ln(\text{Patents}_{f,t}) + \sum_{q \in Q_P} (\beta_q \cdot \text{Patent Quality Share}_{f,q}) + \sum_{q \in Q_I} (\theta_q \cdot \text{Internal Share}_{f,q}) + \epsilon_{f,t},$$

where f and t index firms and five-year periods. The set of patent quality quartiles Q_P are indexed by q and we measure effects relative to the lowest two quality quartiles. For internal patents, we define indicator variables for internal patents being a (0, 20%] share of the firm's total innovation during the period or greater than 20%.

Table 8: Firm-level regression comparison

	Dependent variable is growth to next period		
	Model	Data using citations for quality	Data using claims for quality
Log employment$_t$	-0.0980 (0.0032)	-0.0983 (0.0075)	-0.1012 (0.0076)
Log patents$_t$	0.1091 (0.0048)	0.1310 (0.0125)	0.1330 (0.0125)
Share patents [50, 75)$_t$	0.0894 (0.0150)	0.1004 (0.0379)	-0.0015 (0.0397)
Share patents [75,100]$_t$	0.0734 (0.0135)	0.3659 (0.0399)	0.1274 (0.0382)
(0,1) Medium internal patents$_t$	-0.0579 (0.1105)	-0.0473 (0.0329)	-0.0431 (0.0323)
(0,1) High internal patents$_t$	-0.1056 (0.0085)	-0.1870 (0.0321)	-0.2036 (0.0323)

Notes: See Table 6.

On the whole, the model and data display very similar properties at the micro-level. Firm growth is increasing in total patents, increasing in the share of these patents falling in the upper half of the distribution, and decreasing in the share of the patents that are internal in nature. The data tends to show greater growth effects with patent quality than the model for the very top quartile, but most of the coefficient magnitudes are quite comparable. In the last column, we use patent claims to measure quality and find comparable results.[25]

Appendices C2 and C3 report additional data analyses that confirm features present in the model. C2 shows that the patents that firms develop in their first two years of existence have higher external impact than those subsequently developed by the same firm. C3 shows that the external work that builds on a particular invention tends to have greater forward impact than the internal work that also builds on the same invention. These two features are distinctive elements of our model structure that are important to confirm in the data. Our NBER working paper also provides additional empirical elements that support the model's features. We show, for example, that the external citation distributions that exist for an external patent do not depend upon the size of the firm making the patent. This invariance provides support for our model's structure that

[25] While citations are the more commonly used measure, there is some concern that firm growth or survival could influence future external citations (e.g., out of fear of litigation). We thank a referee for pointing out this feature, which is not directly testable as quality would be observationally similar. Claims provides a check against this concern.

relates firm size to choices over types of innovations, rather than firms of different sizes having inherently different capacities for producing high-quality innovations.

6.4 Extensions

This section considers extensions and robustness checks. We continue to conclude that $\sigma + \psi = 0.9$ is a good estimate for the level of decreasing returns to external innovation in firm size.

6.4.1 Adding Fraction of Top Innovations as a Target

Our model predicts that the fraction of major innovations in a firm's portfolio tends to be decreasing in firm size if external work does not scale one-for-one. This theoretical prediction was empirically verified in Section 5.4, and we used this as an untargeted moment to assess the model. As an alternative exercise, we introduce this empirical moment as an additional target. Table 9 reports the new moments and the new estimate of σ. To save space, the rest of the parameter estimates are not reported.

TABLE 9: ROBUSTNESS WITH FACTS 1 AND 2

Moment	**Data**	**Model**	**Moment**	**Data**	**Model**
profitability	0.109	0.106	entry rate	0.058	0.066
R&D intensity	0.041	0.041	average growth rate	0.010	0.010
internal/external cite	0.774	0.767	growth vs size (fact 1)	-0.035	-0.038
fraction of internal patents	0.215	0.250	top innov. vs size (fact 2)	-0.0034	-0.0034
Estimated σ : 0.395, Implied $\sigma + \psi = 0.895$.					

The model replicates both facts very closely, while also preserving the goodness of fit with the rest of the moments. The resulting estimated σ value is very similar at 0.395.

6.4.2 Adding Patent per Employment as a Target

Table 10 further incorporates the normalized patents per employment regression coefficient as an additional target.

TABLE 10: ROBUSTNESS WITH FACTS 1, 2, AND 3

Moment	**Data**	**Model**	**Moment**	**Data**	**Model**
profitability	0.109	0.113	average growth rate	0.010	0.009
R&D intensity	0.041	0.049	growth vs size (fact 1)	-0.035	-0.057
internal/external cite	0.774	0.806	top innov. vs size (fact 2)	-0.0034	-0.0061
fraction of internal patents	0.215	0.272	patent per emp vs size (fact 3)	-0.182	-0.081
entry rate	0.058	0.059			
Estimated σ : 0.407, Implied $\sigma + \psi = 0.907$.					

While the fit of the first two facts declines with this augmented model, all three relationships are still captured. Most important, the scaling estimate $\sigma = 0.407$ remains robustly identified.

6.4.3 Alternative Growth Cap

The major moment influencing σ in the benchmark estimation in Table 3 is the empirical relationship between firm size and growth. To confirm these results are not sensitive to the winsorization imposed, in Table 11 we keep all parameters at their baseline levels and re-estimate σ with the maximum growth rate of 3000%, versus 1000% in our baseline.

Table 11: Robustness with Growth Rate Maximum

Moment	Data	Model	Moment	Data	Model
profitability	0.109	0.106	entry rate	0.058	0.066
R&D intensity	0.041	0.041	average growth rate	0.010	0.010
internal/external cite	0.774	0.732	growth vs size (fact 1)	-0.048	-0.046
fraction of internal patents	0.215	0.252			
Estimated σ : 0.384, Implied $\sigma + \psi = 0.884$.					

This adjustment lowers σ to 0.384, which is intuitive given that the weaker winsorization allows us to pick up even more abnormal growth for smaller firms, but the influence on our results is overall quite modest.

6.4.4 Alternative R&D Elasticities

Table 12 studies the robustness of our results to alternative estimates of the R&D elasticity, centered on the $\psi = 0.5$ elasticity from the micro studies (see the discussion in Section 6.2.1). Panel A considers a lower value of $\psi = 0.4$, whereas Panel B considers a larger value $\psi = 0.6$.

Table 12: Robustness with Different R&D Elasticities

Panel A. $\psi = 0.4$					
Moment	**Data**	**Model**	**Moment**	**Data**	**Model**
profitability	0.109	0.097	entry rate	0.058	0.067
R&D intensity	0.041	0.041	average growth rate	0.010	0.009
internal/external cite	0.774	0.773	growth vs size (fact 1)	-0.035	-0.036
fraction of internal patents	0.215	0.252			
Estimated σ : 0.497, Implied $\sigma + \psi = 0.897$.					
Panel B. $\psi = 0.6$					
profitability	0.109	0.094	entry rate	0.058	0.068
R&D intensity	0.041	0.039	average growth rate	0.010	0.010
internal/external cite	0.774	0.798	growth vs size (fact 1)	-0.035	-0.036
fraction of internal patents	0.215	0.228			
Estimated σ : 0.283, Implied $\sigma + \psi = 0.883$.					

The model continues to replicate the targeted moments well. The remarkable result is the robustness of the sum of the elasticity parameters $\sigma + \psi \approx 0.9$, which conforms to benchmark estimates.

7 Conclusion

Firms come in many shapes and sizes, as do their innovations. An important step for research on the origins of innovation and endogenous growth is to build an apparatus that can handle more of this firm-level heterogeneity; it is equally important to discern when this apparatus adds value commensurate with its extra complexity. This paper has sought to take a step forward on both of these dimensions. First, we have built a model that allows for internal and external innovation by firms, intuitively links firm-level innovation investments to their sizes using observations from the economics and management literature, and traces out many of the consequences that these differences can yield in terms of firm-level dynamics and aggregate growth rates. The model remains surprisingly tractable with these added ingredients, laying bare some economic factors that can lie behind empirical regularities like deviations from Gibrat's Law or the disproportionate representation of small firms and start-ups among the producers of major innovations.

We also quantified a generalized form of our model using U.S. data from the Census Bureau for 1982-1997. We particularly wanted to assess how significant the elements present in our baseline model constituted a departure from the Klette and Kortum (2004) framework that matches many micro-level facts about firms, without firms making different investment choices. In general, we found the decreasing returns to external innovation in larger firms to be an important but modest departure. For many applications, researchers may want to use the Klette and Kortum (2004) framework directly if their focus is on factors beyond innovation heterogeneity. We have done so in our own work elsewhere, and the results of this paper suggest that these departures can be innocent for many applications. Other researchers may want to build models that directly include R&D scaling differences, and our quantitative estimates can be useful for these efforts. Yet, these differences are meaningful enough that researchers focused on heterogeneous innovation types may want to incorporate elements of our baseline model into their work.

Amongst these contributions, our paper is also quite novel in how it layers on patents and citations across patents to inform the model behavior. Indeed, estimations of our model and the scaling parameters would not have been possible otherwise. This work also allows us to conclude that growth impacts of external innovation have exceeded internal work for the recent U.S. economy, which in turn helps identify some of the special role that small, innovative firms and new entrants can play in economic growth. There is great potential for further developing this link of patents and patent citations and the information they contain into growth models. Our framework is a natural launching point for estimating the role of intellectual property protections for the incentives to innovate and the subsequent trade-offs that come with monopoly rights. As second example, one could follow inventors out of large incumbent firms and into the formation of new companies to study the role of firm spawning in economic growth and the implications of regulations like non-compete clauses. The general take-away from this work is that growth models can garner greater insights and realism by layering information similar to patents and citations that can be studied in both the model and data.

References

Acemoglu, D. (2008): *Introduction to Modern Economic Growth*. Princeton University Press.

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

Acemoglu, D., U. Akcigit, N. Bloom, and W. Kerr (2013): "Innovation, Reallocation, and Growth," NBER Working Paper # 18993.

Acemoglu, D., U. Akcigit, D. Hanley, and W. Kerr (2015): "Transition to Clean Technology," *Journal of Political Economy*, forthcoming.

Acemoglu, D., and D. Cao (2015): "Innovation by Entrants and Incumbents," *Journal of Economic Theory*, 157, 255–294.

Acs, Z. J., and D. B. Audretsch (1987): "Innovation, Market Structure, and Firm Size," *Review of Economics and Statistics*, 69(4), 567–574.

——— (1988): "Innovation in Large and Small Firms: An Empirical Analysis," *American Economic Review*, 78(4), 678–690.

——— (1991): "Innovation and Size at the Firm Level," *Southern Economic Journal*, 57(3).

Aghion, P., C. Harris, P. Howitt, and J. Vickers (2001): "Competition, Imitation and Growth with Step-by-Step Innovation," *Review of Economic Studies*, 68(3), 467–492.

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

Aghion, P., P. Howitt, and J. Vickers (1997): "Competition and Growth with Step-by-Step Innovation: An Example," *European Economic Review*, 41(3), 771–782.

Agrawal, A., I. Cockburn, and C. Rosell (2010): "Not Invented Here? Innovation in Company Towns," *Journal of Urban Economics*, 67(1), 78–89.

Akcigit, U. (2010): "Firm Size, Innovation Dynamics and Growth," University of Pennsylvania Working Paper.

Akcigit, U., H. Alp, and M. Peters (2015): "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," University of Pennsylvania Working Paper.

Akcigit, U., and W. R. Kerr (2010): "Growth through Heterogeneous Innovations," NBER Working Paper # 16443.

Arkolakis, C. (2011): "A Unified Theory of Firm Selection and Growth," NBER Working Paper # 17553.

Balasubramanian, N., and J. Sivadasan (2011): "What Happens When Firms Patent? New Evidence from US Economic Census Data," *Review of Economics and Statistics*, 93(1), 126–146.

Barro, R., and X. Sala-i Martin (1995): *Economic Growth*. New York, NY: McGraw&Hill.

Baumol, W. J. (2002): "Entrepreneurship, Innovation and Growth: The David-Goliath Symbiosis," *Journal of Entrepreneurial Finance*, 7(2), 1–10.

Bernstein, S. (2015): "Does Going Public Affect Innovation?," *Journal of Finance*, forthcoming.

Bloom, N., R. Griffith, and J. Van Reenen (2002): "Do R&D Tax Credits Work? Evidence from a Panel of Countries 1979–1997," *Journal of Public Economics*, 85(1), 1–31.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81(4), 1347–1393.

Blundell, R., R. Griffith, and F. Windmeijer (2002): "Individual Effects and Dynamics in Count Data Models," *Journal of Econometrics*, 108(1), 113–131.

Caballero, R. J., and A. B. Jaffe (1993): "How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in *NBER Macroeconomics Annual 1993, Volume 8*, pp. 15–86. MIT press.

Cabral, L., and J. Mata (2003): "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 93(4), 1075–1090.

Cai, A. J. (2010): "Knowledge Spillovers and Firm Size Heterogeneity," University of New South Wales Working Paper.

Caves, R. E. (1998): "Industrial Organization and New Findings on the Turnover and Mobility of Firms," *Journal of Economic Literature*, 36(4), 1947–1982.

Christensen, C. (1997): *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA: Harvard Business School Press.

Cohen, W. (1995): "Empirical Studies of Innovative Activity," in *Handbook of the Economics of Innovations and Technological Change*, ed. by P. Stoneman. Oxford, UK: Blackwell.

Cohen, W. M., and S. Klepper (1996): "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D," *Review of Economics and Statistics*, 78(2), 232–243.

Davis, S., J. Haltiwanger, and S. Schuh (1996): *Job Creation and Destruction*. Cambridge, MA: MIT Press.

Dunne, T., M. J. Roberts, and L. Samuelson (1988): "Patterns of Firm Entry and Exit in US Manufacturing Industries," *RAND Journal of Economics*, 19(4), 495–515.

Duranton, G. (2007): "Urban Evolutions: The Fast, the Slow, and the Still," *American Economic Review*, 97(1), 197–221.

Duranton, G., and H. G. Overman (2005): "Testing for Localization Using Micro-geographic Data," *Review of Economic Studies*, 72(4), 1077–1106.

Eeckhout, J., and B. Jovanovic (2002): "Knowledge Spillovers and Inequality," *American Economic Review*, 92(5), 1290–1307.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100(3), 1195–1213.

Foster, L., J. Haltiwanger, and C. J. Krizan (2000): "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in *New Developments in Productivity Analysis*. Chicago, IL: University of Chicago Press.

Galasso, A., and T. S. Simcoe (2011): "CEO Overconfidence and Innovation," *Management Science*, 57(8), 1469–1484.

Gans, J. S., D. H. Hsu, and S. Stern (2002): "When Does Start-up Innovation Spur the Gale of Creative Destruction?," *RAND Journal of Economics*, 33(4), 571–586.

Garcia-Macia, D., C.-T. Hsieh, and P. Klenow (2014): "How Destructive is Innovation," Stanford University Working Paper.

Geroski, P. A. (1998): "An Applied Econometrician's View of Large Company Performance," *Review of Industrial Organization*, 13(3), 271–294.

Gilbert, R. J., and D. M. G. Newbery (1982): "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, 72(3), 514–526.

Griliches, Z. (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28, 1661–1707.

——— (1992): "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 94, S29–S47.

Gromb, D., and D. S. Scharfstein (2002): "Entrepreneurship in Equilibrium," NBER Working Paper # 9001.

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

Hall, B., and J. Van Reenen (2000): "How Effective Are Fiscal Incentives for R&D? A Review of the Evidence," *Research Policy*, 29(4), 449–469.

Hall, B. H. (1992): "R&D Tax Policy During the 1980s: Success or Failure?," *Tax Policy and the Economy*, 7, 1–36.

Hall, B. H., A. Jaffe, and M. Trajtenberg (2005): "Market Value and Patent Citations," *RAND Journal of Economics*, 36(1), 16–38.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper # 8498.

Hall, B. H., and R. H. Ziedonis (2001): "The Patent Paradox Revisited: An Empirical Study of Patenting in the US Semiconductor Industry, 1979–1995," *RAND Journal of Economics*, 32(1), 101–128.

Hausman, J., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52(4), 909–38.

Hellmann, T., and E. Perotti (2011): "The Circulation of Ideas in Firms and Markets," *Management Science*, 57(10), 1813–1826.

Henderson, R. M. (1993): "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry," *RAND Journal of Economics*, 24(2), 248–270.

Henderson, R. M., and K. B. Clark (1990): "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly*, 35(1), 9–30.

Hopenhayn, H., G. Llobet, and M. Mitchell (2006): "Rewarding Sequential Innovators: Prizes, Patents, and Buyouts," *Journal of Political Economy*, 114(6), 1041–1068.

Hopenhayn, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), 1127–1150.

Howitt, P. (1999): "Steady Endogenous Growth with Population and R&D Inputs Growing," *Journal of Political Economy*, 107(4), 715–730.

Hsieh, C.-T., and P. Klenow (2014): "The Life-Cycle of Manufacturing Plants in India and Mexico," *Quarterly Journal of Economics*, 129(3), 1035–1084.

Hurst, E., and B. W. Pugsley (2011): "What do Small Businesses Do?," *Brookings Papers on Economic Activity*, 43(2 (Fall)), 73–142.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty (2000): "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review Papers and Proceedings*, 90(2), 215–218.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108(3), 577–598.

Jarmin, R. S., and J. Miranda (2002): "The Longitudinal Business Database," Center for Economic Studies Working Paper.

Jones, C. I. (1995): "R&D-based Models of Economic Growth," *Journal of Political Economy*, 103(4), 759–784.

Jovanovic, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50(3), 649–70.

Jovanovic, B., and G. M. MacDonald (1994): "The Life Cycle of a Competitive Industry," *Journal of Political Economy*, 102(2), 322–347.

Kerr, W., R. Nanda, and M. Rhodes-Kropf (2014): "Entrepreneurship as Experimentation," *Journal of Economic Perspectives*, 28(3), 25–48.

Kerr, W. R. (2010): "Breakthrough Inventions and Migrating Clusters of Innovation," *Journal of Urban Economics*, 67(1), 46–60.

Kerr, W. R., and S. Fu (2008): "The Survey of Industrial R&DPatent Database Link Project," *Journal of Technology Transfer*, 33(2), 173–186.

Klepper, S. (1996): "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *American Economic Review*, 86(3), 562–583.

Klepper, S., and E. Graddy (1990): "The Evolution of New Industries and the Determinants of Market Structure," *RAND Journal of Economics*, 21(1), 27–44.

Klette, T. J., and S. Kortum (2004): "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 112(5), 986–1018.

Kortum, S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65(6), 1389–1420.

Kortum, S., and J. Lerner (2000): "Assessing the Contribution of Venture Capital to Innovation," *RAND Journal of Economics*, 31(4), 674–692.

Kueng, L., M.-J. Yang, and B. Hong (2014): "Sources of Firm Life-Cycle Dynamics: Differentiating Size vs. Age Effects," NBER Working Paper # 20261.

Lamoreaux, N. R., K. L. Sokoloff, and D. Sutthiphisal (2011): "The Reorganization of Inventive Activity in the United States in the Early Twentieth Century," in *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, ed. by N. R. Lamoreaux, and D. Costa, pp. 235–274. University of Chicago Press.

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

——— (2014): "Optimal Growth Through Product Innovation," University of Wisconsin Working Paper.

Lerner, J. (1994): "The Importance of Patent Scope: An Empirical Analysis," *RAND Journal of Economics*, 25(2), 319–333.

——— (1997): "An Empirical Exploration of a Technology Race," *RAND Journal of Economics*, 28(2), 228–247.

——— (2012): *The Architecture of Innovation: The Economics of Creative Organizations*. Boston, MA: Harvard Business School Press.

Lerner, J., M. Sorensen, and P. Strömberg (2011): "Private Equity and Long-run Investment: The Case of Innovation," *Journal of Finance*, 66(2), 445–477.

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9(2), 508–523.

Luttmer, E. G. (2007): "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 122(3), 1103–1144.

——— (2011): "On the Mechanics of Firm Growth," *Review of Economic Studies*, 78(3), 1042–1068.

March, J. G. (1991): "Exploration and Exploitation in Organizational Learning," *Organization science*, 2(1), 71–87.

Mehta, A., M. Rysman, and T. Simcoe (2010): "Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion," *Journal of Applied Econometrics*, 25(7), 1179–1204.

Nelson, R., and S. Winter (1982): *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Nicholas, T. (2014): “Scale and Innovation During Two U.S. Breakthrough Eras,” Harvard Business School Working Paper.

Peretto, P. F. (1998): “Technological Change, Market Rivalry, and the Evolution of the Capitalist Engine of Growth,” *Journal of Economic Growth*, 3(1), 53–80.

Rausch, L. (2010): “Indicators of U.S. Small Business's Role in R&D,” National Science Foundation Info Brief 10-304.

Reinganum, J. F. (1983): “Uncertain Innovation and the Persistence of Monopoly,” *American Economic Review*, 73(4), 741–748.

Romer, P. M. (1986): “Increasing Returns and Long-run Growth,” *Journal of Political Economy*, 94(5), 1002–1037.

Romer, P. M. (1990): “Endogenous Technological Change,” *Journal of Political Economy*, 98(5), S71–102.

Rosen, R. J. (1991): “Research and Development with Asymmetric Firm Sizes,” *RAND Journal of Economics*, 22(3), 411–429.

Samila, S., and O. Sorenson (2011): “Venture Capital, Entrepreneurship and Economic Growth,” *Review of Economics and Statistics*, 93(1), 338–349.

Spence, M. (1984): “Cost Reduction, Competition, and Industry Performance,” *Econometrica*, 52(1), 101–121.

Sutton, J. (1997): “Gibrat's Legacy,” *Journal of Economic Literature*, 35(1), 40–59.

Thomke, S. (2003): *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*. Boston, MA: Harvard Business School Press.

Thompson, P., and M. Fox-Kean (2005): “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment,” *American Economic Review*, 95(1), 450–460.

Trajtenberg, M. (1990): “A Penny for Your Quotes: Patent Citations and the Value of Innovations,” *RAND Journal of Economics*, 21(1), 172–187.

Wilson, D. J. (2009): “Beggar Thy Neighbor? The In-state, Out-of-state, and Aggregate Effects of R&D Tax Credits,” *Review of Economics and Statistics*, 91(2), 431–436.

Zucker, L., M. Darby, and M. Brewer (1998): “Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises,” *American Economic Review*, 88, 290–306.

Appendix

A Proofs of Propositions

Proof of Lemma 1. Note that $Y^* = (1-\beta)^{\frac{1-2\beta}{\beta}} \tilde{\beta}^{\frac{\beta-1}{\beta}} L^* \bar{q}$. Therefore the growth rate of aggregate output is equivalent to the growth rate of the average quality of product lines. We can express the level of $\bar{q}(t)$ after an instant Δt as

$$\bar{q}(t+\Delta t) = \left\{ \begin{array}{c} \bar{q}(t)\left[\tau^* \Delta t (1+\bar{s}) + z^* \Delta t (1+\lambda)\right] \\ +\bar{q}(t)\left[1-\tau^*\Delta t - z^*\Delta t\right] \end{array} \right\}.$$

Now subtract $\bar{q}(t)$ from both sides and divide by Δt and take the limit as $\Delta t \to 0$

$$g = \frac{\dot{\bar{q}}(t)}{\bar{q}(t)} = \lim_{\Delta t \to 0} \frac{\bar{q}(t+\Delta t) - \bar{q}(t)}{\Delta t} \frac{1}{\bar{q}(t)} = \tau^* \bar{s} + z^* \lambda.$$

■

Proof of Proposition 1. Conjecture that

$$V(\mathbf{q}) = A \sum_{q_j \in \mathbf{q}} q_j. \tag{43}$$

Substituting this expression into the original value function,

$$r^* A \sum_{q_j \in \mathbf{q}} q_j = \max_{x, [z_j]_{j \in \mathcal{J}_f}} \left\{ \begin{array}{c} \sum_{q_j \in \mathbf{q}} \pi^* q_j - \sum_{q_j \in \mathbf{q}} \hat{\chi} z_j^{\hat{\psi}} q_j - \Phi \bar{q} \\ -\tilde{\chi} x^{\tilde{\psi}} \bar{q} + x A \bar{q}(1+\bar{s}) \\ + \sum_{q_j \in \mathbf{q}} z_j A q_j \lambda - \sum_{q_j \in \mathbf{q}} \tau^* A q_j \end{array} \right\}.$$

This expression holds if and only if

$$r^* A = \max_z \left\{ \pi^* - \hat{\chi} z^{\hat{\psi}} + z A \lambda - \tau^* A \right\}, \text{ and} \tag{44}$$

$$\max_x \left\{ x A (1+\bar{s}) - \tilde{\chi} x^{\tilde{\psi}} \right\} - \Phi = 0. \tag{45}$$

Assume for now that there is positive entry (we will verify this later in the proof). Then from the free-entry condition (8) we have

$$A = \frac{\nu}{1+\bar{s}}. \tag{46}$$

The maximization in (44) implies $z = \left[\frac{A\lambda}{\hat{\psi}\hat{\chi}}\right]^{\frac{1}{\hat{\psi}-1}}$ or

$$z_j = \left[\frac{\lambda \nu}{(1+\bar{s})\hat{\psi}\hat{\chi}}\right]^{\frac{1}{\hat{\psi}-1}}$$

and

$$\tau = \frac{\pi}{A} + \hat{\chi} \left[\frac{\lambda}{\hat{\psi}\hat{\chi}} \right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} A^{\frac{1}{\hat{\psi}-1}} \left(\hat{\psi} - 1 \right) - g - \rho$$

where the last line used the fact that $r = g + \rho$. Since the growth rate is $g = \tau\bar{s} + z\lambda$, the above expression can be further refined as

$$\tau = \frac{1}{(1+\bar{s})} \left[\frac{\pi}{A} - \left[\frac{\lambda}{\hat{\psi}\hat{\chi}} \right]^{\frac{\hat{\psi}}{\hat{\psi}-1}} A^{\frac{1}{\hat{\psi}-1}} \hat{\chi} - \rho \right].$$

Now we turn to the maximization problem in (45) which delivers the optimal innovation effort (together with (46)) as

$$x = \left[\frac{\nu}{\tilde{\psi}\tilde{\chi}} \right]^{\frac{1}{\tilde{\psi}-1}}.$$

Hence the condition in (45) is

$$\max_{x} \left\{ xA(1+\bar{s}) - \tilde{\chi}x^{\tilde{\psi}} \right\} = \left[\frac{v}{\tilde{\psi}\tilde{\chi}} \right]^{\frac{\tilde{\psi}}{\tilde{\psi}-1}} \tilde{\chi} \left(\tilde{\psi} - 1 \right).$$

Hence assumption 2 guarantees (45). ■

Proof of Proposition 2. Conjecture the form $\mu_n^* = \tilde{A}\tilde{B}^n \frac{1}{n!}$. Then the flow equations in (26) imply

$$F\tilde{A}\tilde{B}^2\tau + x_e = F\tilde{A}\tilde{B}(x+\tau)$$

and

$$\tilde{B}^2\tau = \tilde{B}(x^* + n\tau^*) - nx$$

Combining these two equations implies

$$F\tilde{A}\tilde{B}n\tau^* - F\tilde{A}nx + x_e = F\tilde{A}\tilde{B}\tau$$

This equation can hold for all $n \geq 2$ if and only if $\tilde{B} = x/\tau$ and $\tilde{A} = \frac{x_e}{Fx}$. ■

Proof of Proposition 6. Note that the new value function in general form is

$$rV(\mathbf{q}) - \dot{V}(\mathbf{q}) = \max_{\substack{x_n \in [0,\bar{x}], \\ \{z_j \in [0,\bar{z}]\}_{\mathcal{J}_f}}} \left\{ \begin{array}{c} \sum_{q_j \in \mathbf{q}} \left[\pi^* q_j - \hat{\chi} z_j^{\hat{\psi}} q_j \right] - \bar{q}\tilde{\chi} n^{\tilde{\sigma}} x_n^{\tilde{\psi}} \\ +nx_n \left[\mathbb{E}_j V(\mathbf{q} \cup_+ \{q_j + \bar{q}s_j\}) - V(\mathbf{q}) \right] \\ + \sum_{q_j \in \mathbf{q}} z_j \left[V(\mathbf{q} \backslash_- \{q_j\} \cup_+ \{q_j(1+\lambda)\}) - V(\mathbf{q}) \right] \\ + \sum_{q_j \in \mathbf{q}} \tau \left[V(\mathbf{q} \backslash_- \{q_j\}) - V(\mathbf{q}) \right] \end{array} \right\}$$

Substituting the conjecture $V(\mathbf{q},\bar{q}) = A\sum_{q_j\in\mathbf{q}} q_j + B_n\bar{q}$ into the above value function we get

$$r\sum_{q_j\in\mathbf{q}} Aq_j + rB_n\bar{q} - B_n\bar{q}g = \max_{\substack{x_n\in[0,\bar{x}],\\ \{z_j\in[0,\bar{z}]\}_{\mathcal{J}_f}}} \left\{ \begin{array}{c} \sum_{q_j\in\mathbf{q}} \left[\pi^* q_j - \hat{\chi} z_j^{\hat{\psi}} q_j\right] - \bar{q}\tilde{\chi} n^{\tilde{\sigma}} x_n^{\tilde{\psi}} \\ +nx_n \left[\begin{array}{c} A\bar{q}\left[1+\mathbb{E}_j s_j\right] \\ +B_{n+1}\bar{q} - B_n\bar{q} \end{array} \right] \\ +\sum_{q_j\in\mathbf{q}} z_j A q_j \lambda \\ +\sum_{q_j\in\mathbf{q}} \tau\left[-Aq_j + B_{n-1}\bar{q} - B_n\bar{q}\right] \end{array} \right\}$$

Now equating the terms with q_j and $\bar{q}$ we get

$$rA = \max_{z_j}\left\{\pi^* - \hat{\chi} z_j^{\hat{\psi}} + z_j A\lambda - \tau A\right\}$$

and

$$rB_n - B_n g = \max_{x_n}\left\{ \begin{array}{c} -n^{\tilde{\sigma}}\tilde{\chi} x_n^{\tilde{\psi}} \\ +nx_n\left[A\left[1+\mathbb{E}_j s_j\right] + B_{n+1} - B_n\right] \\ +n\tau\left[B_{n-1} - B_n\right] \end{array}\right\}.$$

Note that from log utility we have $\rho = r - g$. Hence the two value functions become

$$\begin{aligned} rA &= \pi - \tau A + \max_{z_j}\left\{z_j A\lambda - \hat{\chi} z_j^{\hat{\psi}}\right\} \\ \rho B_n &= \max_{x_n}\left\{ \begin{array}{c} -n^{\tilde{\sigma}}\tilde{\chi} x_n^{\tilde{\psi}} \\ +nx_n\left[A\left[1+\bar{s}\right] + B_{n+1} - B_n\right] \\ +n\tau\left[B_{n-1} - B_n\right]. \end{array}\right\} \end{aligned}$$

Now we can take the first order conditions

$$z_j = \left[\frac{A\lambda}{\hat{\psi}\hat{\chi}}\right]^{\frac{1}{\hat{\psi}-1}} \text{ and } x_n = \left[\frac{A\left[1+\bar{s}\right] + B_{n+1} - B_n}{\tilde{\psi} n^{\tilde{\sigma}-1}\tilde{\chi}}\right]^{\frac{1}{\tilde{\psi}-1}}.$$

Hence A is defined by the following equation

$$(r+\tau)A = \pi + A^{\frac{\hat{\psi}}{\hat{\psi}-1}}\left[\frac{\lambda}{\hat{\psi}}\right]^{\frac{\hat{\psi}}{\hat{\psi}-1}}\left(\hat{\psi}-1\right)\hat{\chi}^{\frac{1}{1-\hat{\psi}}}$$

and B_n :

$$B_{n+1} = \left[\frac{(\rho + n\tau)B_n - n\tau B_{n-1}}{\tilde{\psi}-1}\right]^{\frac{\tilde{\psi}-1}{\tilde{\psi}}} \tilde{\psi}\tilde{\chi}^{\frac{1}{\tilde{\psi}}} n^{\frac{\tilde{\sigma}-\tilde{\psi}}{\tilde{\psi}}} + B_n - A\left[1+\bar{s}\right]$$

■

Proof of Proposition 7. First we compute the number of citable patents M. The measure of citable patents after Δt is simply

$$M(t+\Delta t) = \left[M(t)+1\right]\left(x\Delta t(1-\theta) + z\Delta t\right) + 1\times x\Delta t\theta + (1 - x\Delta t - z\Delta t)M(t).$$

Imposing the steady state condition $M(t+\Delta t) = M(t)$ we find $M = \frac{1}{\theta} + \frac{z}{x\theta}$. Recall the flow equations (38) and (39). Equation (38) and (18) imply $\Upsilon_{s_k,0} = \frac{\tau(1-\theta)^k\theta}{M[\tau\theta+\gamma s_k(\tau(1-\theta)+z)]}$. Then we can rewrite (39) in a recursive form as $\Upsilon_{s_k,n} = \Upsilon_{s_k,n-1}\frac{\gamma s_k(\tau(1-\theta)+z)}{[\tau\theta+\gamma s_k(\tau(1-\theta)+z)]}$ which implies $\Upsilon_{s_k,n} = \Upsilon_{s_k,0}\left[\frac{\gamma s_k(\tau(1-\theta)+z)}{\tau\theta+\gamma s_k(\tau(1-\theta)+z)}\right]^n$. Similar reasoning applies to $\Upsilon_{\lambda,n}$ and to the flow equations (40) and (41).

For the second part of the theorem, we just rewrite the same flow equations without the internal citations z. Then the expressions follow. ■

B Full Predictions of Baseline Model

This appendix outlines the full set of predictions for the baseline theoretical model without scaling. Most predictions are general and do not depend upon whether internal or external R&D has a larger average step size. Predictions C3, D5, and D6 are specific to the case of external R&D having the larger step size, which we find empirically to be true. Our NBER working paper provides the proofs of these predictions.

A: Firm Size Distribution and Firm Growth Rates

A1 The size distribution of firms is highly skewed.

A2 The probability of a firm's survival is negatively related to its size.

A3 Small firms that survive tend to grow faster than larger firms. Among larger firms, this negative relationship weakens.

A4 The variance of growth rates is higher for smaller firms.

A5 Younger firms have a higher probability of exiting, but those that survive tend to grow faster than older firms.

B: Firm Size Distribution and Innovation Intensity

B1 R&D expenditures increase with firm size among innovative firms, but the intensity of R&D decreases with firm size.

B2 Similarly, patent counts increase with firm size among innovative firms, but the intensity of patenting decreases with firm size.

B3 Younger firms are more R&D and patent intensive than older firms.

C: Patent Citation Behavior and Innovation Spillover Size

C1 A large fraction of patents receive zero external citations.

C2 The distribution of citations is highly skewed.

C3 An average external patent receives more external citations than an internal patent.

C4 The distribution of patent citation life is highly skewed.

D: Innovation Type and Firm Size Distribution

D1 The proportion of a firm's patents that receives zero future external citations rises with firm size.

D2 The proportion of a firm's backwards citations that are self citations rises with contemporaneous firm size.

D3 Average future external citations per patent is decreasing in firm size.

D4 The relative rate of major innovations (highly cited patents) is higher for small firms. This higher relative rate weakens with more stringent citation quality thresholds.

D5 The average citations (received) of patents by entrants is higher than the average citations of patents by incumbents. Similarly, the average citations of patents by young firms is higher than the average citations of patents by older firms.

D6 The patents made by firms at their entry on average receive more external citations than later patents of the same firm.

E: Innovation Type and Firm Growth Rates

E1 More cited patents lead to higher growth for a firm. This effect is larger for small firms.

E2 An external patent leads to higher growth than an internal patent on average.

E3 More R&D and patent intensive firms grow faster.

E4 Everything else equal, firms that obtain more external patents are more likely to survive. Firms that receive more external citations are more likely to exit the economy.

C Additional Empirical Results

We include here some selected empirical results that provide special details relevant to our model. Our working paper contains a more extensive set of work that we do not repeat here.

C.1 Monte Carlo Simulations of Internal Patent Citations

Table A1 considers in greater detail the observation made in the Introduction that self-citation behavior rises with firm size. We study this issue using patent data and assignees, which allows us to undertake the simulations outside of the Census Bureau. We consider patterns for patents filed in 1995 and their citations over the previous five years. This short period lowers the computation demands of the simulations, and this snap shot is very representative of the general behavior across the full sample. In 1995, the self citation share grows from 9% for firms filing just one patent to 17% for firms filing 2-5 patents. The share further increases to 31% for firms filing over 100 patents.

The last three columns of Table A1 evaluate these observed self citation shares against counterfactuals. Large patenting firms are more likely to cite themselves due to the greater likelihood that they draw upon their past work. This is true even if citations are random. If IBM and a small firm in 1995 draw a random citation for the computer industry from 1990-1995, the likelihood that IBM draws itself is much greater. The likelihood of self citing for a new entrant is naturally zero. This bias to firm size is particularly true where large firms dominate narrow technology fields.

Table A1. Cross-sectional relationship of assignee size and self citation behavior

	Count of assignees by number of 1995 patents with citations for patents over the prior 5 years	Mean observed self citation share for patents over the prior 5 years	Comparison of observed self citation behavior against 1000 Monte Carlo simulations replicating technologies and citation years		
			Mean test statistic for 95% confidence level by size category	Share of firms deviating at 95% confidence level from random behavior	Mean deviations of observed citation shares (col. 2 minus col. 3)
	(1)	(2)	(3)	(4)	(5)
1 patent	8044	9%	1%	13%	8%
2-5 patents	3382	17%	3%	35%	14%
6-10 patents	595	22%	4%	64%	18%
11-20 patents	307	23%	4%	73%	19%
21-100 patents	288	27%	4%	89%	23%
100+ patents	65	31%	6%	97%	25%

Notes: Table reports the results of Monte Carlo simulations of self citation behavior by firm size. The sample is restricted to US-based, industrial patents in 1995 and their citations to other US-based, industrial patents over the prior five years. Rows group assignees by their patent counts in 1995. The second column indicates the share of observed citations that are self citations. For the Monte Carlo simulations, we draw counterfactuals that match the technologies and application years of cited patents. We include the original citation among the possible pool of patents, and we draw with replacement. We measure from the simulation a counterfactual self citation share to assignee size relationship. We repeat the simulations 1000 times to generate 95% confidence bands for the self citation ratio of each assignee. These confidence bands are specific to assignees based upon their size and underlying technologies. The third column provides the mean test statistic by firm size. This statistic rises with firm size because firms with larger patent portfolios are more likely to cite themselves even if citations are random. The fourth column indicates the share of assignees by size category that exhibit self citation behavior that exceeds a random pattern at a 95% confidence level. These deviations are strongly increasing in firm size. The last column presents the mean deviation of observed self citation behavior from the simulation baselines. These deviations are also increasing in firm size.

To confirm that this mechanical effect is not driving the observed relationship in Column 2, we undertake Monte Carlo simulations where we replace observed patents with random counterfactuals. For each observed citation, we draw a counterfactual that matches the technology and application year of the cited patent. We include the original citation among the possible pool of patents, and we draw with replacement. We measure from the simulation a counterfactual self citation share to assignee size relationship. As this relationship depends upon the randomness of the simulation draws, we repeat the procedure 1000 times.

We use these 1000 simulations to generate 95% confidence bands for the self citation ratio of each assignee. These confidence bands are specific to assignees based upon their size and underlying technologies. These confidence bands more rigorously test whether the observed self citation relationships are a systematic departure from the null hypothesis of being randomly determined. As anticipated, Column 3 shows that the mean value of the test statistic is rising in firm size.

Columns 4 and 5 confirm that the observed self citation behavior is a significant departure among large assignees. Column 4 examines the prevalence of departures. For assignees with one patent during 1995, only 13% display self citation behavior that we can reject as being random at a 95% confidence level. This non-random share grows to 97% for assignees with more than 100 patents in 1995. Column 5 also shows that average deviation of self citation shares from the random baseline is growing in firm size. These departures indicate that our results are due to firm behavior rather than the mechanics of firm size. These self citation findings hold in within-firm panel analyses, too.[26]

C.2 Panel Relationship Between Entry and Patent Quality

Table A2 presents some simple panel evidence on patent quality within firms over time. We restrict the sample to new entrants during 1977-1994. We regress traits of patents on an indicator variable

[26]This analysis closely relates to the patent localization work of Jaffe et al. (1993) and Thompson and Fox-Kean (2005). Similar procedures are used in agglomeration calculations like Duranton and Overman (2005) and Ellison et al. (2010). Agrawal et al. (2010) discuss related issues with respect to large patenting firms in "company towns" and their self citation behavior (e.g., Eastman Kodak in Rochester, NY).

for whether or not the patent is filed in the first two years that a firm is observed. We include firm fixed effects to compare early patents of the firm to later patents. We also include technology-year fixed effects. Column 1 shows that the average external citation count is higher at entry. Column 2 shows that patents also have larger numbers of claims at firm entry than in later years. Columns 3-6 show the distribution of external citations in quartiles. Column 3 is the lowest quality quartile, and Column 6 is the highest quality quartile. Entrants have disproportionate representation in the highest quality quartile compared to later years for the same firm. The results describe the time path of firms in terms of invention quality.

Table A2. Panel relationship between entry and patent quality

	Number of external citations	Number of claims on patent	Prevalence of patents by external citation ranks (coefficients sum to zero across columns)			
			0-24%	25-49%	50-74%	75-100%
	(1)	(2)	(3)	(4)	(5)	(6)
First two years the firm is observed	1.1621	0.6920	-0.0148	-0.0042	-0.0048	0.0239
	(0.1557)	(0.1811)	(0.0048)	(0.0059)	(0.0063)	(0.0058)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Technology-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table quantifies changes in average patent quality within firms over time. Columns 1 and 2 show that external citation rates and claims per patent are higher at firm entry. Columns 3-6 show the distribution of external citations in quartiles. Column 3 is the lowest quality quartile, and Column 6 is the highest quality quartile. The coefficients for a row sum to zero across these columns. Entrants have disproportionate representation in the highest quality quartile compared to later years for the same firm. The sample includes 260,972 US industrial patents for firms first observed between 1977 and 1994. Estimations include firm fixed effects and technology-year fixed effects, cluster standard errors at the firm level, and weight patents such that each firm receives constant weight.

C.3 Dynamic Evidence on Quality Within Firms

Table A3 provides evidence to verify our model's assumption that major external innovations are followed within firms by internal innovations and refinements. This process requires that an external innovation be made to dramatically push forward the technology of a product line that is dominated by internal inventions within the currently leading firm. We can further verify these features by demonstrating that the mean quality of citing patents outside of the original firm for a given invention is higher than the mean quality of citing patents within the firm.

We use a linear specification of the form

$$\text{Cite}_{p_2,p_1} = \phi_{p_1} + \eta_{i,t}^{p_2} + \beta \cdot \text{External}_{p_2,p_1} + \epsilon_{p_2,p_1},$$

where $\text{Cite}_{p2,p1}$ models traits of patents p_2 that cite patents p_1. We include citations for U.S. industrial patents filed during 1975-1984. We restrict the citations to be US industrial patents filed within a ten-year window of the original patent. We find similar patterns when using all citations, but the consistent window is more appropriate.

The primary regressor is the indicator variable $\text{External}_{p_2,p_1}$ that takes unit value if the assignee of citing patent p_2 differs from the assignee of cited patent p_1. Three-quarters of citations are external. We include ϕ_{p_1} fixed effects for cited patents. We thus compare differences between internal and external citations on the same patent. We also include $\eta_{i,t}^{p_2}$ fixed effects for the technology i and year t of the citing patent p_2; the patent fixed effects naturally control for these traits for cited patents p_1. We define $\eta_{i,t}^{p_2}$ through USPTO sub-categories and five-year time periods. We cluster standard errors by cited patents.

The first column of Table A3 models the number of external citations on citing patents p_2 as the outcome variable. The second column alternatively tests the number of claims on the citing patent as a measure of quality. Columns 3-6 then test the quality distribution of citing patents in a format similar to Table A2. Quality distributions are determined through ranks of external citations by technology and period. Coefficients across the final four columns for a row approximately sum to zero, but the relationship does not hold exactly given that quality distributions are calculated over a larger group than the regression sample.

Table A3. Assignee size and building upon technologies

	Number of external citations on citing patent	Number of claims on citing patent	Prevalence of patents by external citation ranks among citing patents (coefficients sum to zero across columns)			
			0-24%	25-49%	50-74%	75-100%
	(1)	(2)	(3)	(4)	(5)	(6)
External citation	0.849	1.236	-0.015	-0.009	-0.005	0.029
	(0.053)	(0.073)	(0.002)	(0.002)	(0.002)	(0.002)
Cited patent fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Citing tech-year effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table characterizes differences in patent quality for internal versus external patents that cite a particular invention. Columns 1 and 2 show that external citation rates and claims are higher. Columns 3-6 show the quality distribution of the citations by quartiles. Column 3 is the lowest quality quartile, and Column 6 is the highest quality quartile. External citations are consistently of higher quality. The sample includes 761,940 citations of US industrial patents from 1975-1984 applied for within ten years after the original patent. Estimations include cited patent fixed effects and technology-period fixed effects for citing patents. Estimations cluster standard errors by cited patent.

The first column finds that the mean number of future citations for external work that builds upon a given invention is 0.8 citations higher than the internal work that also builds on the focal invention. This effect is large relative to the sample mean of 8.2. There is also a substantial external premium of 1.2 claims relative to the sample mean of 15.4. Columns 3-6 show that this effect mainly comes from a greater prevalence of upper quartile patents among the external citing patents, with mass moved from the lowest two quartiles of the distribution. These patterns suggest that external work that builds upon a given invention is stronger than the internal work that follows.