Young, Restless and Creative: Openness to Disruption and Creative Innovations*

Daron Acemoglu† Ufuk Akcigit‡ Murat Alp Celik§

November 1, 2017

Abstract

This paper investigates the relationship between openness to disruption and creative innovations—innovations that break new ground in terms of knowledge creation. After presenting a motivating model focusing on the choice between incremental and radical innovation, and on how managers of different ages and human capital are sorted across different firms with different degrees of openness to disruption, we provide firm-level and patent-level evidence consistent with this pattern. Our measures of creative innovations proxy for innovation quality (average number of citations per patent) and creativity (fraction of superstar innovators, the likelihood of a very high number of citations, and generality of patents). Using the age of top managers as a proxy, we present robust evidence that openness to disruption is associated with more creative innovations, but we also show that once the effect of the sorting of young managers to firms that are more open to disruption is factored in, the (causal) impact of manager age on creative innovations is small.

JEL Codes: O40, O43, O33, P10, P16, Z1.

Keywords: corporate culture, creative destruction, creativity, economic growth, entrepreneurship, individualism, innovation, openness to disruption.

*We thank Olga Denislamova, Hyunjin Kim, and Gokhan Oz for excellent research assistance, and conference and seminar participants at the NBER Productivity group, CIFAR, Brown University, Harvard, MIT, University of Pennsylvania, AEA Meetings and the Toulouse Network on Information Technology for helpful suggestions and comments. We also thank Bilkent University Economics Department for the great hospitality during part of this project. Financial support from the National Science Foundation (SES-1459871), the Alfred P. Sloan Foundation (G-2014-13635), the Ewing Marion Kauffman Foundation (20140501), ARO MURI W911NF-12-1-0509 and from the Toulouse Network on Information Technology is gratefully acknowledged.

†Massachusetts Institute of Technology. Email: daron@mit.edu.

‡University of Chicago. Email: uakcigit@uchicago.edu.

§University of Toronto. Email: murat.celik@utoronto.ca.

1 Introduction

We investigate the impact of economic and social incentives on "creative innovations," which we identify with the most influential, innovative and original patents. Though there are currently more than half a million patents granted by the US Patent and Trademark Office (USPTO) per year, only a handful make a fundamental contribution to society's knowledge, and a small fraction account for the bulk of the value created.[1] For example, within the field of drugs and medical inventions, which generated 223,452 patents between 1975 and 2001, the median number of citations within the next five years was four (indicating that only a few other innovations built on them). However, a few patents are much more transformative and also receive many more citations. One was ArthroCare Corporation's 1998 patent for "systems and methods for selective electrosurgical treatment of body structures," which improved a variety of existing surgical procedures and devices used, inter alia, in arthroscopy, neurology, cosmetics, urology, gynecology and laparoscopy/general surgery, and received 50 citations in the next five years. Similarly, in electrical digital data processing, the median number of citations within the next five years in this period is 14, but Sun Microsystems Inc.'s 1994 patent for "method for extracting profiles and topics," which was instrumental in introducing the HTML hypertext system, received 473 citations within the five years.

An idea dating back to Joseph Schumpeter (1934) associates creative innovations and entrepreneurship not only with economic rewards, but also with the ability and desire of potential innovators and entrepreneurs to significantly deviate from existing technologies, practices and rules of organization and society and to engage in "disruptive innovations." This is natural; as Schumpeter emphasizes, innovation is a deviation from existing, inertial ways of doing things, and thus relies on "mental freedom" from, or even "rebellion" against, the status quo (pp. 86-94). Similarly, technologies that will cause the most fundamental "creative destruction" naturally correspond to, and perhaps are driven by, "deviant" and disruptive behavior. This notion is pithily captured by an inscription prominently displayed on the walls of Facebook's headquarters in Silicon Valley: *"Move fast and break things."*

This perspective suggests that organizations and societies that impose a set of rigidly specified rules, discourage initiative and deviations from established norms, shun or even ostracize nonconformist behavior, and do not tolerate those that "move fast and break things" will lag behind their more open or "risk-taking" counterparts in creative innovations—even though they might still be able to function successfully with existing technologies. In the rest of the paper, we thus refer to this constellation of social and economic incentives as *openness to disruption* (short for openness to disruptive innovations, ideas and practices).

[1] See, among others, Trajtenberg (1990), Harhoff et al. (1999) and Sampat and Ziedonis (2004) on the relationship between citations and patent quality.

We first provide a simple model of the interplay between firm type, in particular capturing "*corporate cultures*" varying in the degree to which they are open to disruption and innovation strategies. Firms can engage in an *incremental innovation* by building on their existing leading-edge products. In addition, high-type firms (those that have a corporate culture open to disruption) can attempt a *radical innovation*, which involves combining diverse ideas to generate a technological improvement in a new area. We assume that young managers who have more recently acquired general skills (or are less beholden to a particular type of product or technology) have a comparative advantage in radical innovation, and in consequence, will be hired by firms pursuing radical innovations. In our model, though incremental innovations also increase productivity, it is the radical innovations that are the engine of growth. This is because incremental innovations in a particular "technology cluster" run into diminishing returns (as in Akcigit and Kerr, 2010, or Abrams et al., 2013), while radical innovations create new technology clusters and enable another series of incremental innovations.

Our model predicts a reduced-form cross-sectional relationship between manager age and radical innovation. But this relationship does not correspond to the causal effect of manager age on creative innovations. Rather, manager age is both a determinant of radical innovations and a proxy for a corporate culture that is more open to disruption and/or more conducive to radical innovations; indeed, in our model young managers sort to firms that are open to radical innovation, while also contributing to the likelihood of radical innovations in such firms.[2] These forces can be seen from the longitudinal predictions of the model: firms that hire younger managers should subsequently have more creative innovations (for hiring a young manager is associated either with a change in a firm's type or a change in the firm's innovation strategy as it runs out of productive incremental innovation opportunities). But because firms that are more open to disruption do not immediately switch to radical innovations and hire a young manager, the increase in creative innovations typically proceeds the hiring of a younger manager.

The model clarifies that radical innovations generate higher quality (more highly-cited) patents and tend to be more general in terms of the range of citations they receive (because they are expanding into new areas). This provides us with a strategy to measure the creativity of innovations and investigate the empirical implications of the model.

Our theoretical framework also predicts another relationship we investigate empirically: products with higher sales will encourage high-type firms to pursue incremental innovations which build on their existing product lines, and those with many patents will tilt things in favor of radical innovations (because of diminishing returns).

[2] Both in our theory and empirical work, "openness to disruption" could proxy for other firm characteristics making innovations more radical and creative and simultaneously attracting younger managers, though we emphasize that these are not simply factors affecting the number of innovations or patents, but their composition.

The bulk of our paper is devoted to an empirical study of these ideas. We investigate whether companies with younger managers, which is in part a proxy for greater openness to disruption, engage in more radical and creative innovations. As already noted, manager age is a natural proxy for openness to disruption, since companies with a corporate culture open to disruption are more likely to allow young managers to rise up to the top of the corporate hierarchy.[3]

Our empirical work uses several different measures of creative innovations, all computed from the United States Patents and Trademark Office (USPTO) data. These are *innovation quality*, measured by the average number of citations per patent; *fraction of superstar innovators*, which corresponds to the fraction of patents associated with an innovator classified as a "superstar" on the basis of the number of citations; *tail innovations*, which we measure as the fraction of patents of a company that are at the 99th percentile of the overall citations distribution relative to those that are at the median, thus capturing the likelihood of receiving a very high number of citations normalized by the "median" number of citations; and *generality index*, constructed by Hall, Jaffe and Trajtenberg (2001), which measures the dispersion of the citations that a patent receives from different technology classes.

We establish a robust cross-sectional correlation between CEO age and all of our measures of firm-level creative innovation (with or without a variety of firm-level controls). In summary, firms that tend to employ younger CEOs receive a greater number of citations per patent, have a greater fraction of their patents generated by superstar innovators, have more tail innovations, which are at the very high percentiles of the citations distribution, and have more general patents. We also find similar results when we focus on "within-firm" variation generated by CEO changes: when a younger CEO takes charge, innovations (new patent applications) become more creative. Recall, however, that, as our theoretical analysis highlights, these within-firm results are still a mixture of the sorting effects and the causal effect of manager age on creative innovations (and in fact, consistent with our theory and the importance of sorting, there is also a significant increase in creative innovations *before* a firm switches to a younger CEO).

We then use the structure of our model in conjunction with the reduced-form patterns in the data to shed further light on the relative roles of sorting and the causal effect of manager age on innovation. Namely, we utilize a simple indirect inference procedure to estimate from the reduced-form regression coefficients some of the key parameters of our model, including those governing the causal effect of manager age on creative innovations. This exercise implies that the causal effect of manager age on creative innovations is small relative to the impact of sorting.

We also exploit the patent-level variation to estimate the separate impacts of CEO and inventor

[3]Interestingly, in the examples of major innovations mentioned above, these were produced by companies with unusually young leadership. The average age of top managers at ArthroCare Corporation was 41 at the time, and only 39 at Sun Microsystem (compared to an average age of 54.84 among Compustat companies).

age on the creativity of innovations. Our results indicate that both matter, with roughly similar magnitudes. We find as well that younger CEOs tend to work with younger inventors. These two findings, which suggest that firms typically undertake an assortment of changes when switching towards more radical innovations, further corroborate our interpretation that much of the cross-sectional evidence reflects sorting of younger managers to firms with corporate cultures that are more open to disruption.

Finally, we attempt to shed light on our model's prediction that firms with greater sales and with fewer patents should be less willing to engage in creative innovations by simultaneously including interactions of CEO age with (log) sales and (log) number of patents of the firm in our regressions. The results from this exercise support the notion that CEO age interacts negatively with sales and positively with the number of patents.[4]

Our paper is related to several literatures. First, we build on and extend the literature on the interplay between micro and macro aspects of innovation, in particular Klette and Kortum (2004), by including a choice between radical and incremental innovations, and by incorporating the dimension of matching between managers of different vintages of human capital and type of innovation.[5] Empirical work in this area (e.g., Foster, Haltiwanger and Krizan, 2001, Lentz and Mortensen, 2008, Akcigit and Kerr, 2010, Hurst and Pugsley, 2011, Syverson, 2011, Kogan et al., 2012, Acemoglu et al., 2013) focuses on R&D, patent and productivity dynamics. We depart from this literature both because of our focus on radical (creative) innovations and because we present a detailed analysis of the relationship between creativity of innovations and manager age.

Second, MacDonald and Weisbach (2004), Gorodnichenko and Roland (2012) and Fogli and Veldkamp (2013) are closely related to our work. MacDonald and Weisbach construct an overlapping generations model in which each generation makes technology-specific human capital investments. They show that younger agents are the ones who invest in human capital complementary to new technologies. Their framework does not incorporate innovations and thus has no distinction between radical and incremental innovations. Gorodnichenko and Roland draw a link between innovation and individualism, but focus on aggregate measures of productivity, such as TFP or labor productivity at the country level. In contrast, we start with a microeconomic model of how firms choose their innovation strategies and how managers of different ages endogenously sort across different types of firms, and then exploit firm-level data on the creativity of innovations constructed from patent citations. Fogli and Veldkamp emphasize the role of "individualistic" social networks

[4] The working paper version of our work presented supporting cross-country evidence, showing that the average age of the top managers of the 25 largest listed companies in a country is associated with greater average citations per patent, tail innovations, superstar fraction, and generality of innovations, controlling for the total number of patents, GDP and human capital variables at the country level. This manager age variable is strongly correlated with individualism indices (e.g., from Hofstede, 2001).

[5] This matching aspect is common with theoretical analyses of the role of managers, in particular, Lucas (1978), Garicano (2000), and Garicano and Rossi-Hansberg (2004).

in the diffusion of new technologies, and explore how exposure to different types of diseases is associated with cross-country variation in societal network structures.

Third, our work is linked to the small literature on age and creativity. Galenson and Weinberg (1999, 2001), Weinberg and Galenson (2005), Jones and Weinberg (2011) and Jones (2010) provide evidence that a variety of innovators and top scientists are more creative early in their careers, but they also acquire other types of human capital (perhaps generating different types of creativity) later on. Jones (2009) develops a model in which scientists have to spend more time mastering a given area and have to work in teams because the existing stock of knowledge is growing and thus becoming more difficult to absorb and use.[6]

Fourth, our work is related to the literature pioneered by Bertrand and Schoar (2003) and Bloom and Van Reenen (2007, 2010) which investigates the relationship between CEO and manager characteristics and firm performance. Benmelech and Frydman (2014), for example, show that military CEOs pursue more conservative investment and financial strategies (lower investment in R&D), are less likely to be involved in financial fraud, and perform better during times of distress. Bennedsen et al. (2008) demonstrate that the death of a CEO or shocks to the CEO that potentially affect her focus (death of an immediate family member) impact profitability or operating returns. Also noteworthy in this context is Barker and Mueller (2002), who show that firms with younger CEOs spend more on R&D (though this pattern does not show up in our sample).

The rest of the paper is organized as follows. Section 2 presents our motivating model. Section 3 describes our data. Section 4 presents our main empirical results. Section 5 concludes, while Appendix A contains the proofs from Section 2 and Appendix B, which is not-for-publication, presents additional figures and empirical results.

2 Motivating Theory

In this section, we provide a simple model of radical and incremental innovations to motivate both the conceptual underpinnings of our approach and some of our empirical strategies.

2.1 Production

We consider a continuous-time economy in which discounted preferences are defined over a unique final good as $\int_0^\infty e^{-\rho t} \frac{C(t)^{1-\nu}-1}{1-\nu} dt$,where $\rho > 0$ is the discount rate, $C(t)$ is consumption at time t, and ν is the inverse of the intertemporal elasticity of substitution. The final good is produced using

[6] See also Sarada and Tocoian (2015), who investigate the impact of the age of the founders of a company on subsequent performance using Brazilian data; Azoulay, Manso and Zivin (2011), who document the impact of changes in incentives driven by large academic awards and grants on creativity; and Azoulay, Zivin and Wang (2010), who investigate the impact of the death of a very productive co-author on academic productivity.

There is also an extensive literature in social psychology, mostly using survey and experimental evidence, on age and various attitudes both in general and in business. See, e.g., the survey by Walter and Scheibe (2013).

labor and a continuum of intermediates, each located along a circle, $\mathcal{C}$, of circumference 1, via the constant elasticity of substitution production function,

$$Y(t) = \frac{1}{1-\beta}\left(\int_{\mathcal{C}} q_j(t)^{\beta} k_j(t)^{1-\beta}\, dj\right) L^{\beta}, \tag{1}$$

where $k_j(t)$ denotes the quantity and $q_j(t)$ the quality (productivity) of the leading-edge intermediate j used in final good production at time t, while L is the total amount of production labor, which is supplied inelastically. Consumption is given as $C(t) = Y(t) - K(t)$, where $K(t)$ denotes total spending on intermediates.

We follow Klette and Kortum (2004) in defining a firm as a collection of leading-edge technologies. A perfectly enforced patent on each leading-edge technology is held by a firm, which can produce it at constant marginal cost γ in terms of the final good. Because costs and revenues across intermediates are independent, a firm will choose price and quantity to maximize profits on each of its intermediates (which we also refer to as product lines). In doing so, it will face an iso-elastic inverse demand derived from equation (1), which can be written, suppressing time arguments, as: $p_j = L^{\beta} q_j^{\beta} k_j^{-\beta}, \forall j \in \mathcal{C}$. The profit-maximization problem of the firm with the leading-edge technology for intermediate j can then be written as

$$\Pi(q_j) = \max_{k_j \geq 0}\left\{L^{\beta} q_j^{\beta} k_j^{1-\beta} - \gamma k_j\right\} \ \forall j \in \mathcal{C}.$$

The first-order condition of this maximization problem implies a constant markup over marginal cost, $p_j = \gamma/(1-\beta)$, and thus

$$k_j = \left[(1-\beta)/\gamma\right]^{\frac{1}{\beta}} L q_j. \tag{2}$$

Equilibrium profits for a product line with technology q_j are

$$\begin{aligned} \Pi(q_j) &= \beta\left[(1-\beta)/\gamma\right]^{\frac{1-\beta}{\beta}} L q_j \\ &\equiv \pi q_j, \end{aligned}$$

where the second line defines the constant π.

For future reference, we denote the current period's knowledge stock—current average technology—by

$$\bar{q}_t \equiv \int_{\mathcal{C}} q_{jt}\, dj. \tag{3}$$

2.2 Managers

In addition to workers, the economy is populated by managers, who both play an operational role (reducing costs for firms) and manage innovation.

Managers enter and exit the economy following a stationary Poisson birth and death process, so that the measure of managers, M, and their age distribution is constant over time. We index a manager by her birth date b, or equivalently by her age, $a = t - b$. Denoting the death rate of managers by δ, the fact that the measure of managers is constant at M implies that the age distribution of managers is simply given by an exponential distribution, i.e., the fraction of managers who are below the age a is $1 - e^{-\delta a}$.[7]

A manager acquires the useful knowledge associated with the average technology in the period in which she is born (time b), giving her a knowledge base of

$$\bar{q}_b \equiv \int_{\mathcal{C}} q_{jb} dj.$$

Managers will be hired by monopolists to manage production and innovation on their product lines. In equilibrium, they will be paid a wage $w_{b,t}$ as a function of the current period's average technology, $\bar{q}_t$, and their knowledge, $\bar{q}_b$. We assume that $M < 1$, and this implies that the measure of managers is less than the measure of product lines in the economy, so some product lines will not use a manager. This simplifies the analysis by providing a convenient boundary condition for the determination of equilibrium wages of managers. We also assume that M is not too small, which will ensure that all firms that need a manager for a "radical innovation," as described next, are able to hire one (one can take $M \to 1$ without any loss of generality).

2.3 Corporate Culture and Innovation Dynamics

The economy is populated by two types of firms, with firm type denoted by $\theta \in \{\theta_H, \theta_L\}$ and $\theta_H > \theta_L$. Firm type does not affect productivity directly, but influences the success of radical innovations. In particular, high-type firms ($\theta = \theta_H$) are those with corporate cultures that are open to disruption, and will thus have a comparative advantage in radical innovations. In contrast, we will suppose that low-type firms ($\theta = \theta_L$) are incapable of engaging in radical innovations, which is captured by setting $\theta_L = 0$. Firm type is initially determined upon entry (as described in the next subsection). Thereafter, a low-type firm switches to high type at flow rate $\varphi \in (0, 1)$.[8]

The productivity of each intermediate product is determined by its location along a quality ladder in a given product line. In addition, as noted above and following Klette and Kortum (2004), each leading-edge technology gives the firm an opportunity for further innovation. Innovation dynamics at the firm level are determined by whether the firm pursues an *incremental innovation* or a *radical innovation* strategy.

Incremental Innovation Both types of firms can engage in incremental innovation, which improves the productivity of a product line within the current *technology cluster*. A technology cluster

[7]It is also straightforward to see that denoting the birth rate of managers by δ^{birth}, $M = \delta^{birth}/\delta$.

[8]We assume that there are no switches from high type to low type to simplify the expressions and the analysis.

here refers to a specific family of technologies for that product line. Because incremental innovations take place within this technology cluster, they run into diminishing returns. We model this by assuming that the additional productivity improvements generated by an innovation are decreasing in the number of prior incremental innovations within a technology cluster. Namely, the n^{th} incremental innovation in a technology cluster improves the current productivity of product line j by a step size $\eta_n(q_j, \bar{q}_t)$, where q_j is the current productivity of the technology, $\bar{q}_t$ is the current period's technology, and

$$\eta_n(q_j, \bar{q}_t) = [\kappa \bar{q}_t + (1-\kappa) q_j] \eta \alpha^n \tag{4}$$

with $\alpha \in (0,1)$, $\eta > 0$, and $\kappa \in (0,1)$. This functional form implies two features. First, each innovation builds both on the current productivity of the product line where it originates, with weight $1-\kappa$, and on average technology, $\bar{q}_t$, with weight κ. Second, productivity gains from incremental innovations decrease geometrically, at the rate α, in the number of prior incremental innovations in the technology cluster.

We assume that all firms (regardless of their type) can successfully innovate incrementally at the exogenous rate $\xi > 0$.

Radical Innovations High-type firms can also undertake radical innovations, which combine the current technology of the product line the firm is operating, the knowledge base of the manager, and the available knowledge stock of the economy to innovate in a new area (creatively destroying the leading-edge technology of some other firm). Similar to Weitzman's (1998) approach based on recombination, this combination of knowledge bases creates a new technology cluster.

A radical innovation originating from a particular product line initiates a new technology cluster in a different product line (and the innovating firm will still keep its original product line). The creation of a new technology cluster provides the innovator with the opportunity to start a new series of incremental innovations. Because radical innovations are not directed and each firm controls an infinitesimal fraction of all products, the likelihood that it will be the firm itself radically innovating over its own product is zero.[9] Thus radical innovations are associated with "Schumpeterian creative destruction." We next describe the technology for radical innovations.

A successful radical innovation leads to an improvement over the product line uniformly located on the circle $\mathcal{C}$, and thus generates creative destruction. If there is a successful radical innovation over a product line with technology q_j, this leads to the creation of a new leading-edge technology (now under the control of the innovating firm and manager), with productivity

$$q_j^0 = q_j + \eta_0,$$

[9] It may be more plausible to assume that radical innovations also take place over a range of products that are "technologically close" to the knowledge base of the innovator. Provided that there is a continuum of products within this range, this modification has no impact on any of our results.

where the superscript 0 designates the fact that a radical innovation initiates a new cluster with no prior incremental innovations.

Managers' Role For each of their active product lines, firms hire managers who influence their revenues in two ways. First, a manager of age $a = t - b$ contributes $\bar{q}_t f(a)$ to the revenues of a firm when the aggregate technology level is $\bar{q}_t$ (e.g., by reducing costs).[10] We presume (but do not need to impose) that f is increasing, so that more experienced managers are better at cost reductions. If the firm hires no manager, then it does not receive this additional revenue. Second, a manager affects the flow rate of success for firms attempting a radical innovation as we describe next.

A firm of type θ has a baseline flow rate of radical innovation (regardless of whether it is pursuing radical or incremental innovations) equal to $\psi\Lambda\theta$. In addition, if it pursues a radical innovation strategy, hires a manager with knowledge $\bar{q}_b$ and the current technology in the economy is $\bar{q}_t$, it will generate a flow rate of radical innovation equal to

$$\Lambda\theta\bar{q}^a, \tag{5}$$

where

$$\bar{q}^a \equiv \bar{q}_b/\bar{q}_t$$

is the relative average quality of managers of age a, and $\Lambda \in (0,1]$ (and the superscript, rather than a subscript, here emphasizes that this is a ratio of two averages). This specification confirms that low-type firms, with $\theta_L = 0$, cannot engage in radical innovations—i.e., both $\psi\Lambda\theta_L$ and $\Lambda\theta_L$ are equal to zero.

Since both high- and low-type firms have the same rate of success, ξ, when they attempt incremental innovations, (5) implies that high-type firms and young managers have a *comparative advantage* in radical innovation—only high-type firms can engage in radical innovations, and younger managers contribute to the flow rate of radical innovation with high-type firms.

The parameter Λ captures the role of institutional or social sanctions on radical innovations. Such sanctions may prevent the implementation of certain radical innovations, thus making successful innovations less likely.[11]

Citation Patterns The next example provides more details on the evolution of technology clusters and the citation patterns for the patents related to the incremental and radical innovations.

[10] We model this contribution as an additive element in the revenues of the firm so as not to influence the monopoly price and quantity choices of the firm via this channel.

[11] In the context of our modeling of product lines along the circle $\mathcal{C}$, we may assume that such sanctions permit a firm operating product line j to successfully innovate over technologies that are sufficiently close to itself. Suppose, for example, that j may be allowed to innovate only on product lines that are at most a distance Λ from itself. Then the case of no restrictions would be $\Lambda = 1/2$, so that radical innovations over any product lines on the circle $\mathcal{C}$ are possible, while $\Lambda < 1/2$ would correspond to restrictions and thus lower the likelihood of successful radical innovations.

It illustrates that radical innovations, which create new technology clusters, tend to receive more citations and have greater "generality"—implications we will investigate in our empirical work.

Example 1 The following chart provides an illustrative example focusing on two product lines:

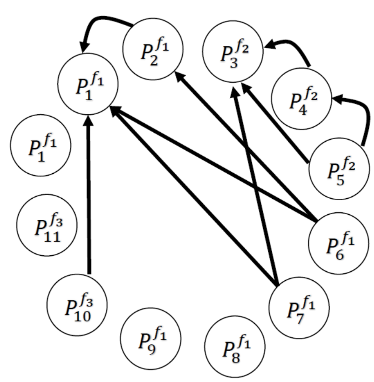
$$\text{First product line:} \quad \Bigg| \; \underbrace{\overbrace{P_1^{f_1}}^{\eta_0}\ \overbrace{P_2^{f_1}}^{\eta_1}\ \overbrace{P_6^{f_1}}^{\eta_2}}_{\textit{Tech Cluster 1}} \; \Bigg| \; \underbrace{\overbrace{P_{10}^{f_3}}^{\eta_0}\ \overbrace{P_{11}^{f_3}}^{\eta_1}}_{\textit{Tech Cluster 2}} \; \Bigg| \; \underbrace{\overbrace{P_{12}^{f_4}}^{\eta_0}\ \overbrace{P_{13}^{f_4}}^{\eta_1}\ \overbrace{P_{14}^{f_4}}^{\eta_2}}_{\textit{Tech Cluster 3}}$$

$$\text{Second product line:} \quad \Bigg| \; \underbrace{\overbrace{P_3^{f_2}}^{\eta_0}\ \overbrace{P_4^{f_2}}^{\eta_1}\ \overbrace{P_5^{f_2}}^{\eta_2}}_{\textit{Tech Cluster 1}} \; \Bigg| \; \underbrace{\overbrace{P_7^{f_1}}^{\eta_0}\ \overbrace{P_8^{f_1}}^{\eta_1}\ \overbrace{P_9^{f_1}}^{\eta_2}}_{\textit{Tech Cluster 2}}$$

Here $P_n^{f_m}$ denotes patent n belonging to firm f_m, and η_n denotes its step size as described in equation (4). In this example, $P_1^{f_1}$, $P_3^{f_2}, P_7^{f_1}$, $P_{10}^{f_3}$ and $P_{12}^{f_4}$ are radical innovations starting new technology clusters and come from high-type firms (f_1, f_2, f_3, and f_4) operating in other product lines. The productivity improvement due to these patents is η_0. Incremental innovations then take place within these technology clusters. For instance, $P_2^{f_1}$, and $P_6^{f_1}$ are incremental innovations in cluster 1 by firm f_1, increasing productivity by step sizes η_1 and $\eta_2 < \eta_1$, respectively.

It is natural to assume that each incremental innovation will cite all previous innovations in its technology cluster, which is the pattern shown in the next figure.

Citation Network



Note: This example illustrates the citations received by patents $P_1^{f_1}, P_2^{f_1}, P_3^{f_2}$, and $P_4^{f_2}$.

In addition, because a radical innovation is recombining ideas from its own product line and the product line on which it is building, it will cite the fundamental ideas encapsulated in the patents that initiated the two technology clusters. For this reason, as shown in the figure below, $P_7^{f_1}$ cites the patents in the technology cluster over which it is innovating, as well as the patents initiating its technology cluster of origin and its destination technology cluster, $P_1^{f_1}$ and $P_3^{f_2}$, respectively, while $P_1^{f_1}$ receives cites from incremental innovations within its technology cluster, from new radical innovations on this product line, and from radical innovations based on this technology taking place in other product lines. As a result, a radical innovation tends to receive more citations as well as more "general" citations; it will also be heavily overrepresented among "tail innovations," meaning among patents receiving the highest number of citations. These are the patterns we will explore in our empirical work.

We close the model by assuming that new firms enter at the exogenous flow rate $x > 0$, and entry corresponds to a (radical) innovation over an existing product line uniformly at random. We further assume that a firm's type is also drawn at random following entry: a successful entrant is high-type, $\theta = \theta_H$, with probability $\zeta \in (0,1)$, and is low-type, $\theta = \theta_L$ $(= 0)$, with the complementary probability, $1 - \zeta$. Thereafter, firm type evolves according to the Markov chain described above.

2.4 The Value of Innovation

Though firms in this economy have a portfolio of product lines and thus the present discounted value of the profits of a firm will depend on this exact portfolio and its evolution, the structure of the equilibrium is greatly simplified because the maximization problem regarding each product is independent of the rest of the portfolio (as in related models such as Klette and Kortum, 2004, and Acemoglu et al., 2013). More formally, let us define $W_s\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)$ (for $s \in \{H, L\}$) as the value of a firm with a vector of products $\overrightarrow{q_f} = \left\{q_{f,j_1}, q_{f,j_2}, ..., q_{f,j_{m_f}}\right\}$, with associated number of incremental innovations $\overrightarrow{n_f} = \left\{n_{f,j_1}, n_{f,j_2}, ..., n_{f,j_{m_f}}\right\}$.[12] Thanks to this independence property, as we show in Appendix A, firm values satisfy

$$W_s\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) = \sum_{m=1}^{m_f} V_s\left(q_j, n\right), \tag{6}$$

where $V_s(q_j, n)$ is the (franchise) value of a product line of productivity q_j with n incremental innovations that belongs to a firm of type $s \in \{H, L\}$ given by

$$\begin{aligned} rV_L(q_j,n) - \dot{V}_L(q_j,n) &= \max_a \left\{\pi q_j + \bar{q}_t f(a) - w_{a,t}\right\} + \xi\left[V_L\left(q_j + \eta_{n+1}, n+1\right) - V_L(q_j,n)\right] \\ &\quad -\tau V_L(q_j,n) + \varphi\left[V_H(q_j,n) - V_L(q_j,n)\right], \end{aligned} \tag{7}$$

and

$$\begin{aligned} rV_H(q_j,n) - \dot{V}_H(q_j,n) &= \max\left\{\begin{array}{c} \pi q_j + \max_a\left\{\bar{q}_t f(a) - w_{a,t} + \xi\left[V_H\left(q_j + \eta_{n+1}, n+1\right) - V_H(q_j,n)\right]\right\}; \\ \pi q_j + \max_a\left\{\bar{q}_t f(a) + \Lambda\theta_H \bar{q}^a \mathbb{E}V_H(\bar{q}_t) - w_{a,t}\right\} \end{array}\right\} \\ &\quad -\tau V_H(q_j,n) + \psi\Lambda\theta_H \mathbb{E}V_H(\bar{q}_t). \end{aligned} \tag{8}$$

Here r is the equilibrium interest rate, τ is the rate of creative destruction in the economy, and $\mathbb{E}V_H(\bar{q}_t)$ denotes the expected value of a radical innovation when the aggregate technology level is $\bar{q}_t$. The form of these value functions is intuitive and instructive about the workings of the model. In (7), a low-type firm's value from a product with productivity q_j that has previously had n incremental innovations depends on the flow of profits, $\pi q_j + \bar{q}_t f(a) - w_{a,t}$, and the additional value

[12] Here and elsewhere, we suppress time as an explicit argument of the value functions to simplify notation.

from an incremental innovation, which arrives at the flow rate ξ and increases the productivity of the firm with step size η_{n+1}. In addition, the second line of (7) captures the fact that the value from this product will disappear at the rate of creative destruction in the economy, τ, representing the replacement of this product by a higher-quality one, and the value of the firm may increase because it may transition to high type at the flow rate φ. A high-type firm's value in (8) is similar except that it involves an additional choice between incremental innovation and radical innovation as represented by the inner maximization.

2.5 Stationary Equilibrium With $\kappa = 1$

We now characterize the stationary equilibrium of this economy. We start with the case where $\kappa = 1$ in equation (4)—so that all current innovations build on current technology, $\bar{q}_t$ (and not on the current productivity of the existing technology cluster). This assumption considerably simplifies the analysis, and we return to the general case where $\kappa < 1$ below.

Characterizing the Stationary Equilibrium A *stationary equilibrium* is defined as an equilibrium in which aggregate output, Y_t, grows at a constant rate g, and the distribution of product lines between high- and low-type firms and over the prior number of incremental innovations remains stationary.

As noted above, firms decide the age of the manager to hire for each of the product lines they are operating and whether to engage in a radical or incremental innovation. Since some firms will not hire managers (as $M < 1$), all firms not attempting a radical innovation on a product line must be indifferent between hiring and not hiring a manager for that product line, which implies that the equilibrium wage for managers, employed by firms engaged in incremental innovations, satisfies the boundary condition:

$$w_{a,t} = \bar{q}_t f(a) . \tag{9}$$

We next turn to the value of a product line operated by a high-type firm, (8). Because of the comparative advantage of young managers for radical innovation in (5), there will exist a maximum age a^* such that only managers below this age will work in firms attempting radical innovation. Moreover, the maximization over the age of the manager in (8) implies that firms engaged in radical innovation must be indifferent between hiring any manager younger than a^*, and thus

$$\bar{q}_t f(a^*) + \Lambda\theta_H \bar{q}^{a^*} \mathbb{E}V_H(\bar{q}_t) - w_{a^*,t} = \bar{q}_t f(a) + \Lambda\theta_H \bar{q}^{a} \mathbb{E}V_H(\bar{q}_t) - w_{a,t} \text{ for all } a < a^*.$$

The boundary condition, (9), implies that the oldest manager hired for radical innovation earns $w_{a^*,t} = \bar{q}_t f(a^*)$. Hence

$$w_{a,t} = \begin{cases} \bar{q}_t f(a) & \text{for} \quad a \geq a^* \\ \bar{q}_t f(a) + \Lambda\theta_H \left[\bar{q}^{a} - \bar{q}^{a*}\right] \mathbb{E}V_H(\bar{q}_t) & \text{for} \quad a < a^* \end{cases} . \tag{10}$$

This wage schedule highlights that younger or older managers might be paid more (this will depend on the f function): younger managers have a comparative advantage in radical innovation, but older managers might be more productive in operating firms.[13]

The next proposition provides the characterization of the stationary equilibrium.

Proposition 1 *Low-type firms ($\theta = \theta_L$) always hire "old" managers (those with $a > a^*$ or $b < b_t^*$), pursue incremental innovations and never generate radical innovations.*

Let λ_n denote the probability that a high-type firm ($\theta = \theta_H$) pursues a radical innovation on a product line with n incremental innovations. There exists an integer n^* such that:

- *high-type firms pursue incremental innovations, $\lambda_n = 0$, on product lines with $n < n^*$ prior incremental innovations and hire "old" managers (those with $a > a^*$ or $b < b_t^*$);*
- *they pursue radical innovations, $\lambda_n = 1$, on product lines with $n > n^*$; and*
- *they pursue radical innovations with probability $\lambda_{n^*} \in [0,1]$ on product lines with $n = n^*$.*

Whenever they pursue radical innovations, high-type firms hire "young" managers (those with $a \leq a^*$ or $b \geq b_t^*$).

A lower Λ, corresponding to the society being less permissive to radical innovations, will increase n^* (so that a lower fraction of high-type firms will pursue radical innovation), and will reduce the wages of young managers (because there is less demand for the knowledge of young managers).

In addition to providing the expression for the threshold n^* and proving this proposition, Appendix A also characterizes (and establishes the existence of) a stationary equilibrium in this economy. This proposition's implications are discussed in the next subsection. Here we simply note that, because the threshold for switching to radical innovation, n^*, is an integer, equilibrium aggregates are not continuous in parameters, and hence the equilibrium may involve some degree of mixing as captured by the fact that $\lambda_{n^*} \in [0,1]$.

Empirical Implications Our empirical work is inspired by Proposition 1. As explained above, radical innovations will be associated with greater indices of our measures of creative innovations. We will first investigate the cross-sectional relationship between manager (CEO) age and creative innovations. In these cross-sectional regressions, manager age is taken to be a proxy of a corporate culture that is more open to disruption. Therefore, from Proposition 1, we expect a negative cross-sectional relationship between manager age and creative innovations. As just stressed, this cross-sectional relationship does not correspond to the "causal effect" of manager age on creativity

[13] The evidence in Galenson and Weinberg (1999, 2001), Weinberg and Galenson (2005) and Jones and Weinberg (2011) is consistent with the possibility that either younger or older creative workers might be more productive.

of innovations (which would apply if we varied manager age holding the firm's corporate culture constant); in particular, it also reflects the sorting of younger managers to corporate cultures that are open to disruption and thus more conducive to creative innovations.

Our model's longitudinal implications—that is, implications about how manager age and creativity of innovations vary over time for a firm—shed further light on the relative magnitudes of the sorting and the causal effects. To understand these implications, let us consider the innovation dynamics of firms implied by Proposition 1. Low-type firms always engage in incremental innovations and never generate radical innovations. High-type firms may attempt a radical innovation depending on how many prior incremental innovations they have had on a product line.

- For a product line with $n < n^*$, a high-type firm hires an old manager (or keeps its already existing old manager) and pursues an incremental innovation strategy. Given the technology specified above, such a firm still generates radical innovations at the rate $\psi\Lambda\theta_H$.

- For a product line with $n \geq n^*$, a high-type firm hires a young manager and engages in radical innovation (with probability λ_{n^*} for $n = n^*$). In this case, the average rate of radical innovation across product lines where radical innovation strategies are pursued can be computed using the fact that the age distribution of managers is given by the exponential distribution, as

$$\psi\Lambda\theta_H + \frac{1}{F(a^*)}\int_0^{a^*} \Lambda\theta_H \bar{q}^a dF(a) = \psi\Lambda\theta_H + \frac{\Lambda\theta_H\delta}{g+\delta}\frac{\left[1 - e^{-(g+\delta)a^*}\right]}{\left[1 - e^{-\delta a^*}\right]}. \tag{11}$$

Now consider a low-type firm that switches to high-type, and to simplify the discussion, suppose that it has a unique product line. Then, if this product line has had $n < n^*$ incremental innovations, the firm will continue to pursue an incremental innovation strategy $\lambda_n = 0$, and keep its old manager.[14] In the process, it will generate radical/creative innovations at the flow rate $\psi\Lambda\theta_H$. When it reaches $n = n^*$, it will hire a young manager, switch to a radical innovation strategy with probability $\lambda_{n^*} > 0$, and at that point, its rate of radical/creative innovations will increase, on average, from $\psi\Lambda\theta_H$ to the expression in (11) times λ_{n^*}. In contrast, if the product line of the firm at the time of switching to high-type has had $n > n^*$ incremental innovations, it will immediately hire a young manager, switch to a radical innovation strategy, and generate radical innovations at the flow rate given by (11).

This discussion clarifies that when we focus on the relationship between within-firm changes in manager age and creative innovations, we expect to find two regularities. First, when a firm switches from an older to a younger manager, this should be associated with an increase in creative innovations. Second, firms that switch from an older to a younger manager should, on average,

[14] Strictly speaking, this is true under an infinitesimal cost of replacing managers. Otherwise, the firm could fire its old manager and hire another old manager, with no impact on our results or discussion here.

experience an increase in creative innovations even *before* the switch. We emphasize that even this further increase following the switch to a younger manager does not correspond to the causal effect of manager age on creative innovations for two reasons; first, these firms will be simultaneously switching to radical innovation and hiring a young manager; and second, the increase in the likelihood of radical innovation will depend on the exact number of prior incremental innovations and the age of the manager hired. For this reason, in Section 4.3 below, we estimate the causal effect by keeping the type of the firm and the number of prior incremental innovations constant, and just changing manager age by a given amount.

Finally, though we will not be able to investigate this directly in our empirical work, the implications of changes in Λ are interesting. A lower value of this parameter naturally reduces radical innovations and, at the same time, decreases the wages of young managers, thus making it look like the society is discriminating against the young; but in fact this is a consequence of the society discouraging radical innovations.

2.6 Equilibrium With $\kappa < 1$

In this subsection, we turn to the general case with $\kappa < 1$. The structure of the equilibrium is similar to the case with $\kappa = 1$, except that now the switch to radical innovation for high-type firms will depend both on their current productivity and on their prior incremental innovations.

Proposition 2 *Consider the economy with $\kappa < 1$. Then, for a product line with current quality q operated by a high-type firm, the manager will be younger and will pursue radical innovation when the number of prior incremental innovations is greater than or equal to $n_t^*(q)$, where $n_t^*(q)$ is increasing in q. That is, a high-type firm is more likely to pursue radical innovation when its current productivity is lower and the number of its prior innovations in the same cluster is higher.*

This proposition thus establishes that in this generalized setup (with $\kappa < 1$), the main results from Proposition 1 continue to hold, but in addition we obtain the new result that radical innovation is more likely when a high-type firm has lower current productivity (conditional on its prior number of incremental innovations); or conversely, for a given level of productivity, it is more likely when there has been a greater number of prior incremental innovations. Intuitively, when the baseline productivity of a product line is higher, the benefits of incremental innovations building on it are also greater and a high-type firm will pursue such incremental innovations for longer before switching to radical innovation. We will investigate this additional implication in our empirical analysis as well.[15]

[15] This result is related to the idea of "disruptive innovations" proposed in Christensen's *The Innovator's Dilemma* (1997). Our result clarifies that our potential answer to the innovator's dilemma, consistent both with Arrow's (1962) replacement effect and the results presented below, is that successful firms with higher sales have more to fear from disruptive innovations, and tend to retrench and become less open to creative innovations.

2.7 Discussion of Assumptions and Microfoundations

In this subsection, we discuss the role and possible microfoundations of the critical assumption underpinning the assignment of young managers to high-type firms and to radical innovation—the comparative advantage in equation (5).

Endogenizing human capital decisions: Our key justification for (5) is that agents acquire the knowledge available at the time they are born. Though this was imposed as a technological feature, it can be readily endogenized (as in Chari and Hopenhayn, 1993, or MacDonald and Weisbach, 2004). The most natural assumption here would be that agents decide when to go to school, and an agent who goes to school for some prespecified period of time, say an interval of length $\Delta > 0$, and graduates at time t acquires the frontier knowledge at that time, q_t as given in (3). Given the stationary structure of the problem, we can make two observations. First, it is always optimal for an agent to acquire schooling immediately (rather than wait and do so at a later date).[16] Second, we can also derive a straightforward sufficient condition ensuring that an agent would never want to go back to school after this initial schooling interval. In particular, once again starting in stationary equilibrium, if a manager of age a at time t does not go back to school, she will have a discounted lifetime income of

$$\int_0^\infty e^{-(r+\delta)s}[\bar{q}_{t+s} f\left(a+s\right) + \max\left\{\Lambda\theta_H\left[\bar{q}^{a+s} - \bar{q}^{a^*}\right], 0\right\}\mathbb{E}V_H(\bar{q}_{t+s})]ds,$$

while if she goes back to school, her discounted lifetime income will be

$$\int_\Delta^\infty e^{-(r+\delta)s}[\bar{q}_{t+s} f\left(a+s\right) + \max\left\{\Lambda\theta_H\left[\bar{q}^{s} - \bar{q}^{a^*}\right], 0\right\}\mathbb{E}V_H(\bar{q}_{t+s})]ds.$$

The latter expression thus enables the agent to reduce $\bar{q}^a$ and potentially earn more from being assigned to high-type firms.[17] However, its comparison to the previous expression makes it clear that if $f(a)$ and Δ are sufficiently large, then it will not be beneficial for a manager to go back to school. For example, an upper bound for the discounted lifetime income from schooling is

$$\int_\Delta^\infty e^{-(r+\delta)s}\bar{q}_{t+s} f\left(a+s\right) + \Lambda\theta_H\left[1 - \bar{q}^{a^*}\right]\mathbb{E}V_H(\bar{q}_{t+s})]ds = \int_\Delta^\infty e^{-(r+\delta)s}\bar{q}_{t+s} f\left(a+s\right)ds + \frac{\bar{q}_t\Lambda\theta_H\left[1-\bar{q}^{a^*}\right]}{r+\delta-g}e^{-\Delta(r+\delta-g)},$$

which assumes that after re-schooling the manager has the highest contribution to innovation forever (whereas in reality her contribution would decline as she ages). On the other hand, the minimum lifetime incomes she would obtain without going to school can be written as

[16] This is because the problem facing an agent at any two dates is identical given the stationary environment and the constant probability of death, δ, and thus if the agent wanted to wait between time t and t', then she would also want to wait indefinitely, violating the transversality condition.

[17] Notice that in writing these expressions, we are interpreting a literally as age, so that when the manager goes back to school, her age is not affected. Alternatively, a could stand for the manager's experience in a particular line of business, and in that case, a could also be reset when she goes back to schooling, which would introduce an additional opportunity cost of returning to school. This does not have an important effect on the qualitative argument here, though it may provide a better approximation to some applications.

$$\int_0^{\Delta} e^{-(r+\delta)s}[\bar{q}_{t+s} \inf f(a)]ds + \int_{\Delta}^{\infty} e^{-(r+\delta)s}\bar{q}_{t+s} f(a+s)\, ds = \int_{\Delta}^{\infty} e^{-(r+\delta)s}\bar{q}_{t+s} f(a+s)\, ds + \frac{1-e^{-(r+\delta-g)\Delta}}{r+\delta-g}\bar{q}_t f_{\min},$$

where $f_{\min} = \inf f(a)$. By comparing these two expressions and noting that their first terms are identical, we obtain a sufficient condition for any manager to never prefer to go back to school,

$$\Lambda\theta_H \left[1-\bar{q}^{a^*}\right] < e^{\Delta(r+\delta-g)} f_{\min}.$$

As already anticipated, this condition is satisfied when Δ or when $f_{\min}$ are large.

An alternative form of comparative advantage: We introduced the comparative advantage of young managers in radical innovation in the simplest possible form—by assuming that they have the same productivity in incremental innovation and greater productivity in radical innovation. Similar results would obtain as well if they have comparable productivity in radical innovation but *lower productivity* in incremental innovation.

Suppose, for example, that all managers have the same rate of arrival of radical innovations given by $\Lambda\theta_H$ when they are employed by high-type firms, but the productivity of a manager aged a in incremental innovation is $\xi g(a)$, where $g(a)$ is increasing. In this case, the pattern of assignment will be slightly different—it will be first the older managers who are assigned to management, but there will still exist a critical age threshold, say a^{**}, such that managers younger than this age will be assigned to high-type firms wishing to specialize in radical innovation. Young managers will also earn strictly less than older managers, but radical innovations will continue to increase following a switch from older to younger managers.

Comparative advantage from competing uses of time: Relatedly, in our baseline model, radical innovations and the operational duties of a manager do not crowd each other out. An alternative, equally natural assumption is that, because seeking radical innovations is time-consuming, it will interfere with the cost-reducing activities of the manager. Under the natural and common assumption that all of these tasks have to be performed by a single manager (i.e., it is not possible to add a separate manager for innovations), attempting radical innovations will have the opportunity cost of reducing the other beneficial roles of the manager. Since experienced managers are more productive at cost reduction and other operational roles, this reasoning directly implies that it will be younger managers who have an effective comparative advantage in radical innovations, even if they are less productive in both operations and radical innovations than older managers.

A re-combinatorial microfoundation for comparative advantage: Another microfoundation for this pattern of comparative advantage is to assume that radical innovation requires recombining different ideas, while more experienced managers will have an expertise in exploiting a specific set of well-established ideas (perhaps ideas with which they have worked before). Such a microfoundation can be developed in a way that generates the pattern of comparative advantage in our baseline model.

One advantage of this alternative line is that the reason why more experienced managers are better at operations, but not as good as young managers in radical innovations, can be endogenized. Specifically, managers may choose to invest in their ability to understand and exploit certain technologies as they age, but this could be at the expense of remaining on top of other ideas, while younger managers may be "jacks of all trades, masters of none," making them less effective in running an established business, and as a result, giving them a comparative advantage in radical innovation.

An organizational microfoundation for comparative advantage: Yet another possibility leading to the pattern of matching in our model would come not from an intrinsic comparative advantage of young managers for radical innovation, but from the potential conflict of interest between managers and owners. Suppose that attempting radical innovation is more costly for managers, and it is difficult for the owners of the firm to verify that the manager is truly attempting radical innovations. This sort of situation will create a major conflict of interest, whereby all managers might wish to claim that they are attempting radical innovations, but in reality may shirk and go for the easy life. If, as it seems plausible, more experienced managers are better able to control the flow of information out of an organization and thus hide their true activities, it might be more difficult for owners to induce these experienced managers to engage in radical innovation. It may then be cheaper and more effective to turn to more "pliable" younger managers when there will be a switch to radical innovation.

Finite lives and risk-taking: The comparative advantage of younger managers in more radical innovations may also come from their greater willingness to take risks, which could in turn have biological roots or may be a consequence of the fact that, when lives are finite, they will have longer horizon than older managers and thus tend to have greater tolerance for risk.

Managers and inventors: We have so far abstracted from inventors, which play an important role both in practice and in our data analysis below. A final alternative structure which leads to similar results is to assume that it is not young managers who are important for radical innovation, but young inventors (a pattern for which we also find support in the data). But if young inventors work better in a team with young managers, for example because older managers would not communicate well with them or would attempt to block some of their ideas, there will again be a pattern in which young managers are assigned to firms specializing in radical innovations.

3 Data and Variable Construction

In this section, we describe the various datasets we use and our data construction. We also provide some basic descriptive statistics.

3.1 Data Sources

USPTO Utility Patents Grant Data (PDP) The patent grant data are obtained from the NBER Patent Database Project (PDP) and contain data for all 3,279,509 utility patents granted between the years 1976-2006 by the USPTO. This dataset includes extensive information on each granted patent, including the unique patent number, a unique identifier for the assignee, the nationality of the assignee, the technology class, and backward and forward citations in the sample up to 2006. Following a dynamic assignment procedure, we link this dataset to the Compustat dataset, which we next describe.[18]

Compustat North American Fundamentals We draw our main sample from the Compustat for publicly traded firms in North America. This dataset contains balance sheets reported by the companies annually between 1974 and 2006. It comprises 29,378 different companies, and 390,467 *company* $\times$ *year* observations. The main variables of interest are net sales, employment, firm age (defined as time since entry into the Compustat sample), SIC code, R&D expenditures, total liabilities, net income, and plant property and equipment as a proxy for physical capital.

Executive Compensation Data (Execucomp) Standard and Poor's Execucomp provides information on the age of the top executives of a company starting from 1992. We use information on CEO age or the average age of (top) managers of a company to construct proxies for openness to disruption at the firm level.[19]

The Careers and Co-Authorship Networks of U.S. Patent Inventors Extensive information on the inventors of patents granted in the U.S. between years 1975-2008 is obtained from Lai et. al.'s (2009) dataset. These authors use inventor names and addresses as well as patent characteristics to generate unique inventor identifiers upon which we heavily draw. Their dataset contains 8,031,908 observations at the *patent* $\times$ *inventor* level, and 2,229,219 unique inventors, and can be linked to the PDP dataset using the unique patent number assigned by the USPTO.

3.2 Variable Construction

Innovation Quality Our baseline measure of innovation quality is the number of citations a patent received as of 2006. We use the truncation correction weights devised by Hall, Jaffe, and Trajtenberg (2001) to correct for systematic citation differences across different technology classes

[18] Details on the assignment procedure are provided at https://sites.google.com/site/patentdataproject/.

[19] We drop observations where reported CEO age is less than 26.

and for the fact that earlier patents will have more years during which they can receive citations. The average innovation quality of a company in a year is computed as the average number of citations of patents the company applied for in that year.

Superstar Fraction A superstar inventor is defined as an inventor who surpasses his or her peers in the quality of patents generated as observed in the sample. A score for each unique inventor is generated by calculating the average quality of all the patents in which the inventor took part. All inventors are ranked according to this score, and the top 5% are considered to be superstar inventors. The superstar fraction of a company in a year is calculated as the fraction of patents with superstar inventors in that year (if a patent has more than one inventor, it gets a fractional superstar designation equal to the ratio of superstar inventors to the total number of inventors).

Tail Innovations The tail innovation index is defined as the fraction of a firm's patents that receive more than a certain number of citations. Namely, let $s_{ft}(p)$ denote the fraction of a firm's patents that are above the p^{th} percentile of the year t distribution according to citations. Our baseline tail innovation index, $\text{Tail}_{ft}(p)$, is simply $s_{ft}(0.99)/s_{ft}(0)$, and thus measures the fraction of patents by firm f at time t with citations above the $99th$ percentile. As an alternative measure, we also consider $\text{Tail}_{ft}(p) = s_{ft}(p)/s_{ft}(0.50)$, where $p > 0.50$. By including $s_{ft}(0.50)$ in the denominator, this alternative measure focuses on whether controlling for their "average" innovation output, some companies generate innovations with very high citations. We also consider our baseline index with $p = 0.90$ as yet another alternative measure for robustness.

Generality and Originality We also use the generality and originality indices devised by Hall, Jaffe and Trajtenberg (2001). Let $i \in I$ denote a technology class and $s_{ij} \in [0, 1]$ denote the share of citations that patent j receives from patents in technology class i (of course with $\sum_{i \in I} s_{ij} = 1$). Then for a patent j with positive citations, we define: $\text{Generality}_j = 1 - \sum_{i \in I} s_{ij}^2$. This index thus measures the dispersion of the citations received by a patent in terms of the technology classes of citing patents. Greater dispersion of citations is interpreted as a sign of greater generality. The originality index is defined similarly except that we use the citations that a patent gives to other patents. The patent classes used are the 80 two-digit International Patent Classification (IPC) classes.

3.3 Sample and Descriptive Statistics

Our baseline analysis focuses on an unbalanced *firm sample* comprised of 7111 observations from 1256 firms between 1992 and 2004. Our data on CEO age does not extend before 1992, and we cannot go further than 2004, since our patent citation data end in 2006 and we need at least two post-grant years for citation analysis. We also study a larger *patent sample* of 316,516 patents (from 1192 distinct firms).

Panel A of Table 1 provides descriptive statistics for our firm and patent samples. Since we focus on regressions weighted by the number of patents held by a company, all statistics are weighted by the number of patents as well. We multiply our indices for tail innovation, superstar fraction, generality, and R&D intensity by 100 for ease of inspection.

The table shows that average CEO age is 55.3 in our firm sample and 55.7 in our patent sample. Panels B and C show that our main measures of creativity of innovations are highly correlated, except for the generality index, which is negatively correlated with our tail innovation index and weakly correlated with the others.

4 Empirical Results

In our theory manager age is partly an indicator of a corporate culture that is open to disruption (because high-type firms that have a comparative advantage in radical innovation select younger managers), but also has a causal effect on creative innovations (since a young manager has a comparative advantage in radical/creative innovations). Motivated by these theoretical results, we start with the cross-sectional relationship between firm-level measures of creative innovations and manager age.[20] We then turn to a more direct investigation of the effect of manager age on creative innovations, focusing on "within-firm" variation. Finally, exploiting the timing of the increase in creative innovations following a change in manager age, we provide estimates of the structural parameters of the model.

4.1 Cross-Sectional Results

Our cross-sectional results are presented in Table $2-4$ and in Figure 1. Our estimating equation is

$$y_f = \alpha m_f + \mathbf{X}_f' \boldsymbol{\beta} + \delta_{i(f)} + \nu_t + \varepsilon_f, \tag{12}$$

where y_f is one of our measures of creative innovations introduced in the previous section (innovation quality, superstar fraction, tail innovation, or generality) for firm f, and m_f is our firm-level measure of openness to disruption—the average age of company CEOs over our sample window. In addition, $\mathbf{X}_f$ is a vector of controls, in this case, firm age, log of employment, log of sales, and log of total number of patents during our time window. Controlling for firm age is particularly important in order to distinguish the correlation of creativity of innovations with manager age from its correlation with firm age. In addition, $\delta_{i(f)}$ denotes a full set of four-digit main SIC dummies, so that the comparisons are always across firms within a fairly narrow industry,[21] and ν_t denotes a full set of

[20] Another caveat is that our theoretical results relate manager age at the product-line level to the innovation strategy and creativity of innovations, while the bulk of our empirical analysis in this section will be at the firm level focusing on the age of a firm's CEO (or top managers).

[21] All firms in our baseline sample are in one of 283 four-digit SIC industries.

year dummies. Finally, ε_f is the error term. To start with, we do not include firm fixed effects, and for this reason we use average age of CEOs in the specifications and focus on the cross-sectional correlation rather than the year-to-year variation in CEO age and our outcome measures.[22] We turn to specifications with fixed effects in Table 5.

Unless otherwise indicated, all of our regressions have one observation per firm×year, and are weighted with the total patent count of the firm in that year, so that they put less weight on observations for which our measures of creative innovations are computed from only a few patents. All standard errors are clustered at the firm level and are robust against heteroscedasticity.

Different columns of Table 2 correspond to our four different measures of creative innovations. Column 1 shows an economically sizable correlation between CEO age and our measure of innovation quality (average number of citations per patent). The coefficient estimate, −0.168 (standard error = 0.075), is statistically significant at less than 5% and indicates that companies with a younger CEO have greater innovation quality. We interpret this pattern as evidence that companies that are more open to disruption (and willing to hire younger managers) tend to be the ones producing more creative innovations. The quantitative magnitudes are sizable but plausible. For example, a one-year decrease in CEO age is associated with a 0.168 increase in average citations, which is approximately 1.1% of the firm-level weighted mean of our innovation quality variable (15.9).

The estimated effects of the covariates are also interesting. Firm age is negatively associated with innovation quality, suggesting that younger firms are more creative (though this pattern is not as robust as the impact of CEO age in other specifications). In addition, our measures of creative innovations are uncorrelated with employment and sales, and are only weakly correlated with the number of patents held by the firm (except for tail innovations). This confirms that our measures of creativity of innovations are quite distinct from the total number of patents.

Column 2 shows a similar relationship with the superstar fraction (−0.319, standard error = 0.133). This result suggests that younger CEOs tend to work with higher-quality innovators (a relationship we directly investigate in Table 10 below). Columns 3 and 4 show even more precisely estimated relationships with our measures of tail innovations and generality. The implied quantitative magnitudes are also a little larger—a one-year increase in CEO age is associated with, respectively, 3.2% and 4.6% increases relative to weighted sample means in these two measures.

The patterns in the data underlying the results in Table 2 are depicted in Figure 1, which plots the correlation between our four measures of the creativity of innovations and CEO age. To transparently illustrate these relationships, for each of our measures, we create deviations from the industry×year means (and group all observations with CEO age ≤ 40 or CEO age ≥ 60).

[22] Panel A of Table 5 shows very similar, but more precisely estimated, results when the CEO age variable is time-varying (without fixed effects).

The negative relationship between CEO age and our four measures is evident. Moreover, these empirical relationships are well approximated by linear regressions (the fitted lines correspond to a linear regression using these industry×year deviations from means on CEO age bins without any other covariates). We show the same relationships when controlling for the same covariates as in Table 2 in Figure B1 in Appendix B.

Tables 3 and 4 probe the robustness of our baseline cross-sectional results in different directions, and demonstrate that under most reasonable variations, the relationship between manager age and the creativity of innovations is, if anything, even stronger than in Table 2.

Table 3 looks at several different specifications. Panel A shows that the results are similar, even if a little smaller, when the regression is unweighted. Panel B estimates our baseline specification using a median regression, which is less sensitive to outliers. The results are again very similar, and in fact typically more precisely estimated, confirming that our baseline results are not driven by outliers. Panel C, which is the regression analogue of Figure 1, shows a very similar relationship when we include only the four-digit SIC dummies and year effects. Panels D and E replace the four-digit SIC dummies in our baseline specification with two-digit and three-digit SIC dummies (192 and 59 dummies, respectively). Panel F goes in the opposite direction and enriches the set of controls; in addition to the baseline covariates in Table 2, it includes several other firm-level controls: profitability (income to sales ratio), debt to sales ratio, and log physical capital of the firm. The results are virtually the same as those in Table 2, but a little more precisely estimated. Panel G, additionally, adds R&D intensity (R&D to sales ratio) to the previous specification.[23] This is intended to verify that our results cannot be explained by some firms performing more R&D than others (here the sample declines to 5,907 observations). The estimates are once again very close to those in our baseline regressions in Table 2, and the R&D intensity variable itself is not significant in any of the columns.

Our baseline regressions are only for firm×year observations with positive patents (since our measures cannot be computed when the denominator is zero). Panel H verifies that this potential endogenous selection into our sample is not responsible for our results. It includes all available firm×year observations, imputing a value of zero to all of our measures when a firm does not have any patents in that year. The estimated relationship between CEO age and our measures of the creativity of innovations are remarkably similar to the baseline results in this case (with only the tail innovation index experiencing a sizable decline in coefficient, which still remains statistically significant at 5%).

Finally, Panel I takes a simple approach to deal with the issue of self-citations (whereby a firm cites its own patents) and includes the fraction of self-citations of the patents of the firm as an

[23] To deal with outliers in R&D expenditures, we winsorize this variable at its 99th percentile value.

additional control variable in our baseline specification. This has little effect on the relationship between CEO age and the creativity of innovations, though this self-citation measure itself is significantly positive, suggesting that firms that give more citations to their own patents tend to be more creative according to all four of our measures.

In addition, Table B1 in Appendix B shows that the results are very similar in a much smaller balanced sample of 279 firms with no missing values for any of our main variables between 1995 and 2000, and also when we use the average age of the top management team rather than CEO age. (We prefer CEO age as our baseline measure because across companies there is considerable variation in the number of managers for which age data are available, making this measure potentially less comparable across firms). It demonstrates as well that the results are very similar in high-tech and low-tech subsamples (where high-tech firms are those in SIC 35 and 36, which include industrial and commercial machinery and equipment and computer equipment; and electronic and other electrical equipment and components, and low-tech firms are the rest), and in a non-pharmaceutical subsample. These subsamples are further studied in Table 6.

Table 4 shows the weighted and the unweighted relationship between CEO age and several alternative measures of creative innovations. These are: a measure of innovation quality using average citations per patent computed using only five years of citations data, a measure of superstar inventors using information on the most highly cited patent of the inventor, the tail innovation index with $p = 0.90$, the alternative tail innovation index we introduced above, which includes the fraction of patents with cites above the median in the denominator, and the originality index. We also look at employment growth, sales growth and the R&D intensity of the firm in that year to both investigate whether the more creative innovations translate into faster growth and to check whether CEO age impacts firm outcomes beyond patents. The results using the alternative measures of the creativity of innovations are similar to those in Table 2, except that the relationship is marginally insignificant at 5% with the alternative measure of superstar fraction when we look at weighted regressions (but are comfortably significant at 5% in the unweighted case). The estimates are also insignificant for the alternative measure of innovation quality and originality in the unweighted regressions. We further find negative effects of CEO age on employment and sales growth in the unweighted regressions, but not in weighted regression, perhaps reflecting the fact that the effects of younger CEOs can be more easily detected on smaller firms. Finally, consistent with our earlier emphasis, we do not find any relationship between CEO age and R&D intensity.

Overall, these results suggest that there is a robust and strong statistical relationship between the age of the CEO and each one of our four measures of creative innovations.

4.2 Panel Results with Firm Fixed Effects

We now show that a strong and fairly robust association between CEO age and creative innovations is present when we focus on within-firm variation in the age of the CEO. We further document that consistent with our theory, creative innovations start increasing before there is a decline in CEO age.

Panel A allows CEO age to vary across years, but still without fixed effects. Thus relative to Table 2, the only difference here is that we are exploiting both the between-firm and over-time variation. As a result, there is now a stronger negative relationship between CEO age and our measures of creativity of innovations than in Table 2.

Panel B turns to our main specification, which includes firm fixed effects as well as year effects (and, of course, in this case, SIC industry dummies and firm age are dropped). This means that we are now focusing on within-firm variation and the CEO age variable is being identified from changes in CEOs—that is, from whether a firm that switches to a younger CEO tends to have more creative innovations relative to its mean.[24] In addition to throwing away all of the between-firm variation, another challenge to finding meaningful results in this specification is that patent applications in one year are often the result of research and product selection from several previous years.[25] These concerns notwithstanding, all of the coefficient estimates on CEO age in these within-firm regressions, except for generality, are negative and statistically significant. For innovation quality, the magnitude of the estimate is about 12% larger than the specification without fixed effects in Table 2 (e.g., −0.188 vs. −0.168), whereas for superstar fraction and tail innovations, it is smaller—about 47% to 62% of the magnitude in Table 2.

The current CEO influences the contemporaneous innovation strategy, and in our model, this has an immediate impact on radical innovations. In practice, some of the impact is likely to be delayed, since research projects, and even patenting, can take several years. We may therefore expect the impact of the CEO's human capital, decisions and age to influence the creativity of innovations over time. We investigate this issue in Panel C by including current CEO age and lagged CEO age simultaneously. Our results show that, with all of our measures of creative innovations (except generality), both matter with quantitatively similar magnitudes.

A related question concerns separating the impact of the current CEO from the persistent effects of past innovations—for example, past creativity may spill over into current creativity in part because patents from a research project may arrive in the course of several years. We investigate this issue by including the lagged dependent variable on the right-hand side. Though such a model,

[24] This specification is related to Bertrand and Schoar's famous (2003) paper on the effect of managers on corporate policies, though in contrast to our focus on CEOs, their sample includes chief financial and operating officers as well as lower-level executives.

[25] Recall, however, that patents are classified according to their year of *application*, so we are investigating the impact of CEO age not on patents granted but on patents applied for when the CEO is in charge.

with fixed effects and lagged dependent variable, is not consistently estimated by the standard within estimator when the coefficient on the lagged dependent variable is close to 1, the results in Panel D show that its coefficient is very far from 1 and the estimates are fairly similar to those in Panel A.[26]

Finally, in Panel E we turn to a central longitudinal implications of our model—that creativity of innovations should increase, on average, before the firm switches to a younger manager. The most direct way of investigating this prediction is by including the lead of CEO age together with current CEO age (similar to the specification in Panel C, except that lead CEO age replaces lagged CEO age). The specifications reported in Panel E show statistically significant negative effects of both current and lead CEO age on the creativity of innovations (except with the generality measure). Interestingly, and perhaps somewhat surprisingly, the magnitudes of the lead and the contemporaneous effects are quite similar. The significant effect of lead CEO age is prima facie evidence of the importance of sorting of younger CEOs to firms that are more open to disruption, an issue we investigate in greater detail in the next subsection.

Table 6 investigates the robustness of within-firm relationships reported in Table 5, focusing on the specifications reported in Panels B and E, and on our first three measures, innovation quality, superstar fraction and tail innovation (since there is no robust relationship with generality once firm fixed effects are included). Panel A documents a very similar relationship when the regression is unweighted. Panel B shows that the results with firm fixed effects are also robust to imputing a value of zero to our measures of creative innovations when the firm does not have any patents in that year (as in Panel H of Table 3). Panels C and D show that the same relationships are present in both our high-tech and low-tech subsamples, with if anything a stronger relationship in the low-tech subsample. Finally, Panel E shows that the results are robust to including the self-citation fraction on the right-hand side as in Panel I of Table 3.[27]

Table B4 in Appendix B further investigates the role of self-citations. Even though the cross-sectional relationships between CEO age and innovation quality or tail innovations disappears when all self-citations are excluded, the results with fixed effects continue to be precisely estimated (there

[26] If we estimate these models using Arellano and Bond's (1991) GMM estimators, the results are similar with innovation quality and superstar fraction, but weaker with the tail innovation index, partly because we lose about a quarter of our sample with these GMM models.

[27] Table B2 shows that our results with firm fixed effects are similar when we include the additional controls from Panel F of Table 3, when we include R&D intensity, and in the non-pharmaceutical sample. We also report a median regression with similar results, though in this case, since the median regression is nonlinear and would not consistently remove the fixed effects, we first de-mean the data. We do not report a specification with average manager age, since the year-to-year variation in this variable is largely driven by the number of managers for whom ages reported.

Finally, Table B3 shows the within-firm equivalent of Table 4, where the relationship between CEO age and the alternative measures of creative innovations continues to be robust. However, the relationship with employment and sales growth is weaker and not statistically significant, which is unsurprising since the impact of a CEO on sales and employment growth is likely realized over time and reflects not just his or her influence on creative innovations but a whole range of other factors.

is no natural analogue of the superstar fraction variable that excludes self-citations).[28] Table B5 turns to the issue of citations added by examiners. Unfortunately, this breakdown of citations is available only after 2002 in our data, leaving us with a time window of three years, 2002-2004, and our baseline results with or without fixed effects are no longer significant in this shorter sample as shown in this table. Nevertheless, when examiner citations (with or without self-citations) are excluded, the relationship between CEO age and innovation quality is similar, but just like the baseline results, is no longer statistically significant.

Overall, the results in this subsection demonstrate that firms that switch to younger CEOs generate more creative innovations both after and shortly before such a switch. Though this pattern is indicative of the simultaneous presence of sorting and causal effects of CEO age on creative innovations, as explained previously it does not directly translate into causal estimates. We next turn to an indirect inference procedure exploiting the structure of our model to obtain such causal estimates.

4.3 The Causal Effect of Manager Age on Creative Innovations

In this subsection, we perform an indirect inference exercise in order to shed further light on the causal effect of manager age on creative innovations. We choose the parameters of the model presented in Section 2 so that the model quantitatively matches the reduced-form estimates—in particular, the coefficients of lead and current CEO age for innovation quality. We then use these implied parameters to compute the implied causal effect of manager age on creative innovations.

The (average) impact of a younger manager on the creativity of innovations for a given firm type is $\frac{1}{F(a^*)}\int_0^{a^*} \Lambda\theta_H \bar{q}^a dF(a) = \frac{\Lambda\theta_H\delta}{g+\delta}\frac{\left[1-e^{-(g+\delta)a^*}\right]}{\left[1-e^{-\delta a^*}\right]}$. Because of the sorting of younger managers to firms that are more open to disruption, we cannot directly obtain this quantity from our reduced-form empirical exercise. Rather, we need to obtain estimates of the parameters ψ and $\Lambda\theta_H$ (the parameters Λ and θ_H do not matter separately, and thus in what follows, we will treat $\Lambda\theta_H$ as a single parameter). The reduced-form coefficient estimates are functions of these parameters, but they also depend on the transitions between high-type and low-type firms, the distribution of incremental innovations per product relative to the threshold for radical innovation, n^*, and the stationary distributions theoretically characterized in Appendix A.

Though structurally estimating all of the underlying parameters of our model would require more information on firm transitions and stationary distributions, we can obtain estimates of the structural parameters that are relevant for the causal effect of CEO age on creative innovations more straightforwardly. For this exercise, we set the discount rate to $\rho = 0.02$, and normalize the

[28] We suspect that the weaker cross-sectional relationship without self-citations reflects the inadequacy of the truncation correction weights we are using, which are devised for overall citations and do not take into account that self-citations typically arrive much earlier than other citations (the mean arrival time of self-citations is about 3.5 years, while for citations from others it is about 8.5 years.

profit flow to $\pi = 1$ (which is without loss of any generality). We fit an exponential distribution to the age distribution of managers in our sample to obtain an estimate of δ in the model. We take the entry rate to be $x = 5\%$ which corresponds to the entry rate in our Compustat sample. Finally, we take the parameter α, which determines how rapidly the productivity of incremental innovations declines from Akcigit and Kerr (2015), who estimate a similar parameter from the patent citation distribution.

This leaves the following parameter vector $\Psi \equiv \{\psi, \varphi, \Lambda\theta_H, \xi, \eta, \zeta\}$ to be determined. Once these parameter values have also been fixed, optimal innovation decisions and equilibrium stationary distributions can be computed using the expressions provided in Section 2 and Appendix A). We can then generate simulated firm histories from which the equivalents of the reduced-form regression coefficients in Table 5 can be computed. Of particular importance for this exercise are the specifications in Panels C and F of Table 5, where various measures of creative innovations were regressed on current CEO age (and lead CEO age in Panel F), firm fixed effects and controls. Throughout we focus on the innovation quality measure (column 1).

Let us denote the coefficient estimate on current CEO age in column 1, Panel C of Table 5 by $\gamma_{current}$, and the coefficient estimates on current and lead CEO age in column 1, Panel F, respectively, by $\gamma'_{current}$ and γ'_{lead}. In our indirect inference procedure, we will target these three parameters. Specifically, we generate data from the model given a parameter vector Ψ, and convert the measure of successful radical innovation in the model, which is a 0-1 variable, into the same units as our innovation quality variable (by dividing it by its variance and multiplying it with the variance of innovation quality). We then run the same regressions as in Panel C and F of Table 5, and compare the estimates to the empirical estimates of $\gamma_{current}$, $\gamma'_{current}$ and γ'_{lead}.

In addition to these three regression coefficients, our indirect inference procedure targets three central moments in the data: the average annual growth rate of (real) sales per worker, within-firm coefficient of variation of radical innovations, and the fraction of incremental innovations, measured as fraction of internal patents (which mainly build on innovating firms' existing lines as opposed to innovating on product lines operated by other firms).[29] This implies that we have in total six data moments and six parameters.

Finally, we make two additional assumptions in matching the model to data. First, in the model managers are employed at the product line level, whereas in the data we only observe CEO at the level of the company (which comprises several product lines). We ignore this distinction, and treat the data as if it were generated from one product firms. Second, in the model, the identity of the manager is indeterminate as there are no costs of changing managers, so a firm could change its manager every instant or at some regular interval even without changing its innovation strategy.

[29] Following Akcigit and Kerr (2015), we define internal patents as those whose majority of citations are self cites.

To make the model more comparable to data, we assume that a firm keeps its manager until it needs to switch from an older to a younger manager in order to change its innovation strategy.

Table 7 provides the values of the parameters we have selected on the basis of external data as well as the values of the parameters in the vector Ψ, which are chosen to match the six aforementioned moments. Table 8 shows the match between the values of these moments in the data and those implied by the model. The model-implied numbers are very close to the targeted empirical moments, and in particular, the model's predictions are consistent with reduced-form regression results, including the significant and sizable coefficient on lead CEO age, which is generated by the fact that $\psi > 0$ and is a non-trivial source of creative innovations.

The implied pattern is also visible in Figures 2 and 3, which plot the probability of a creative innovation and the average CEO age as a function of time since switching to high-type. These figures show that firms slowly reduce the average age of their managers after switching to high-type (since if at first they are below n^*, they do not need to change their CEO). Correspondingly, they also slowly increase their probability of creative innovations. Because much of this increase in the probability of creative innovations takes place before high-type firms switch to a younger manager, in the reduced-form regressions, it will be captured by lead CEO age.

It is also useful to gauge whether, at these estimated parameter values, the model performs well on some other dimensions. One empirical moment we have not used for estimation is the probability of firms switching to younger managers. At the estimated parameter values, 20% of all firms attempt a radical innovation (these are the high-type firms with $n \geq n^*$). Consequently, "young" managers (defined as those with $a < a^*$ in Proposition 1) also make up 20% of the population of managers, implying that a^* corresponds to age 50 in our sample of managers/CEOs. Using this information, we can then compare the annual probability of a firm switching from an old manager (with $a > a^*$) to a young manager (with $a \leq a^*$) in the data and in the model. Reassuringly, these two numbers are fairly close to each other, respectively 4.10% and 2.75%.

Using these parameter estimates, we next compute the "causal effect" of manager age on creative innovations. We focus on the "average treatment effect," corresponding to the impact of having a younger manager for an average firm in our sample. We start with the equilibrium stationary distribution and reshuffle managers across the entire set of firms randomly for 13 periods. We assume that after the reshuffling, each firm will pursue the same innovation strategy as before.[30] We then use the data generated by this thought experiment to run a regression of the likelihood

[30] It is possible that some firms would switch their innovation strategy because they end up with much older or much younger managers. However, whether this is the case or not would also depend on managerial wages after reshuffling, which in turn depends on a variety of auxiliary assumptions on wage determination under "mismatch". Our strategy avoids this complication by estimating a lower bound on this effect, though this lower bound is likely to be fairly tight since low-type firms cannot change their innovation strategy and most high-type firms would be unlikely to alter their innovation strategy either unless there is a very large change in the age of the manager assigned to them.

of a successful radical innovation on manager age for this entire sample of firms (including both low-type firms and high-type firms not attempting a radical innovation). This exercise yields an average causal effect of -0.003. We compare this estimate to the reduced-form relationship between these two variables in the model-generated data, -0.189 (as reported in Table 7). Therefore, this average causal effect accounts approximately for 1.6% ($\simeq 0.003/0.189$) of the relationship between CEO age and creative innovations. The rest of this relationship is explained by sorting effects—i.e., because it is high-type firms that are hiring younger managers.[31]

Overall, our indirect inference exercise establishes that the model can generate the patterns we see in the data, but implies that much of the reduced-form relationship between manager age and creative innovations is due to sorting. Nonetheless, there is a small causal effect of younger managers on creative innovations as well.

4.4 Inventor Age and Creativity of Innovations

We next turn to patent-level regressions to investigate the relationship between the age of inventor—defined as any inventor listed in our patent data—and our various measures of creativity of innovations. Though in our theoretical model there is no distinction between managers and inventors, this distinction is important in practice. One might then expect the role of product-line managers in our model to be played partly by the top management of the firm and partly by inventors (or the lead inventor) working on a particular R&D project. CEOs, then, not only decide which projects the company should focus on but also choose the research team. In this subsection, we bring in information on the age of inventors in order to investigate the simultaneous effects of inventor and CEO age on the creativity of innovations.

We use Lai et. al.'s (2009) unique inventor identifiers described above to create a proxy for this variable. Our proxy is the number of years since the first innovation of the inventor, which we will refer to as "inventor age."

Our main regression in this subsection will be at the patent level and take the form

$$y_{ift} = \phi I_{ift} + \alpha m_{ft} + \mathbf{X}'_{ift}\boldsymbol{\beta} + \delta_f + \gamma_i + d_t + \varepsilon_{ift}. \tag{13}$$

[31] As an alternative, we also computed the analogue of "treatment effect on the treated" in the model by considering the loss of creative innovation that firms that were previously hiring young managers and pursuing radical innovations would suffer. For this exercise, we start with the equilibrium stationary distribution and reshuffle managers only among firms attempting radical innovation (which are high-types with more than n^* incremental innovations and hiring managers younger than a^*), and we repeat this for 13 periods. Because such firms will continue to attempt radical innovation after the reshuffling (since the reshuffling involves only managers younger than a^*), the change in the likelihood of successful radical innovation of any given firm captures the causal effect of a younger (or older) manager on a firm attempting radical innovation. The resulting causal effect is estimated as -0.016, corresponding to about 8% of the reduced-form regression coefficient of -0.189. Yet, the implied magnitude is even smaller, since the -0.189 number is for the entire sample, whereas -0.016 is only for the 20% subsample attempting radical innovation. If we correct for this difference by multiplying the 8% number by 0.20, the resulting estimate of the causal effect is essentially identical to the one we found from our main exercise, about 1.6% of the total effect.

Here y_{ift} is one of our measures of the creativity of innovation for (patent) i granted to firm f at time t. Our key right-hand-side variable is I_{ift}, the age of the inventors named in patent i (in practice, there is often more than one such inventor listed for a patent). In addition, m_{ft} is defined as CEO age at time t and will be included in some regressions, $\mathbf{X}_{ift}$ is a vector of possible controls, and δ_f denotes a full set of firm fixed effects, so that our specifications here exploit differences in the creativity of innovations of a single firm as a function of the characteristics of the innovators involved in the relevant patent. In our core specifications, we also control for a set of dummies, denoted by γ_i, related to inventor characteristics as we describe below. All specifications further control for a full set of year effects, denoted by d_t, and ε_{ift} is the error term.[32]

The results from the estimation of (13) are reported in Table 9. In Panel A we focus on a specification similar to the regressions with firm fixed effects reported in Table 5. This is useful for showing that this different frame still replicates the results showing the impact of CEO age on creativity of innovations. In particular, Panel A focuses on Compustat firms for the period $1992 - 2004$ and includes the same set of controls as in Table 5 Panel A (firm fixed effects, year fixed effects, log employment, log sales and log patents of the firm); it does not contain any variables related to inventor characteristics. As in the rest of this table, these regressions are not weighted (since they are at the patent level), and the standard errors are robust and clustered at the firm level.

Our results using this specification are similar to those of Panel A of Table 5, though a little smaller. In column 1, for instance, we see an estimate of -0.119 (standard error $= 0.038$) compared to -0.188 in Table 5. As a natural patent-level analogue of our tail innovation index, we look at a dummy for the patent in question being above the pth percentile of the citation distribution, and report results using this measure for two values, $p = 0.99$ and $p = 0.90$, in columns 2 and 3. Both of these measures are strongly negatively correlated with CEO age.[33]

Panel B goes in the other direction and reports the estimates of a model that controls for inventor characteristics and looks at the impact of inventor age, without controlling for CEO age, for the same sample as in Panel A (thus restricting it to firms with information on CEO age). As with all of the other models reported in this table, in Panel B we control for a full set of dummies for the (count of) maximum number of patents over our sample period of any of the inventors on this patent;[34] a full set of dummies for the size of the inventor team (i.e., how many inventors are

[32] A single patent can appear multiple times in our sample if it belongs to multiple firms, but this is very rare and applies to less than 0.2% of the patents in our sample.

[33] Our measure of superstar fraction is no longer meaningful at the patent level. For completeness, we also show results with the generality index, even though the results in Table 5 already indicated that, with firm fixed effects included, there is no longer a significant relationship between CEO age and the generality index, and this lack of relationship persists for all of the estimates we report in Table 9.

[34] In other words, we include a dummy variable for the assignee/inventor of this patent with the highest number of total patents having $k = 1, 2, ..., 89+$ patents (where 89+ corresponds to 89 or more patents for the inventor with the maximum number of patents).

listed); and a full set of dummies for the three-digit IPC class.[35] The inclusion of this rich set of dummy variables enables us to compare inventors of similar productivity. It thus approximates a model that includes a full set of inventor dummies.[36] The results show that there is a strong relationship between inventor age and the creativity of innovations. For example, in column 1, the coefficient estimate on inventor age is -0.234 (standard error = 0.026), about twice as large as the CEO age estimate in Panel A.

When we do not control for CEO age, the sample can be extended beyond $1992-2004$. This is done in Panel C, which expands the sample in two different ways, first by including Compustat firms without CEO information, and second by broadening the time period covered to $1985-2004$. The results are very similar to those in Panel B, indicating that the focus on Compustat firms with CEO age information is not responsible for the broad patterns we are documenting.

Panel D extends the sample further to non-Compustat firms, which can also be included in our analysis since we are not using information on CEO age. This increases our sample sixfold (since most patents are held by non-Compustat firms). However, in this case, we can no longer include the employment and sales controls. Despite the addition of almost 1.5 million additional patents and the lack of our employment and sales controls, the results in this panel are again very similar to those in previous panels, and suggest that, at least in this instance, our results are not driven by our focus on the Compustat sample.

Panel E provides our main results in this subsection. It returns to the Compustat sample over the period 1992-2004 and adds back the CEO age variable; otherwise, the specification is identical to that in Panel B. The results show precisely estimated impacts of both CEO age and inventor age. For example, in column 1 with our innovation quality variable, the coefficient on CEO age is -0.119 (standard error = 0.036) and that on inventor age is -0.233 (standard error = 0.026); these are very close to the estimates in Panels A and B, respectively. The pattern is similar in the other columns (except again for generality).

These results provide further evidence that the relationship between manager/CEO age and the creativity of innovations in the data reflects an important dimension of sorting. In particular, firms appear to make several associated changes—in top management and innovation teams—around the same time they change their portfolio of innovation and their innovation strategy (and perhaps their "corporate culture"). Reflecting this sorting, the estimated magnitudes linking CEO age to our indices of creative innovations are smaller in Table 9 than those in our baseline firm-level regressions. Our next results, reported in Table 10, provide some direct evidence on this by looking at the relationship between inventor age and CEO age. We estimate a regression similar

[35] This corresponds to 374 separate technology classes and is roughly at the same level of disaggregation as the SIC dummies we used in the firm-level analysis in Tables 2-3.

[36] We cannot include a full set of inventor fixed effects directly because inventor age would not be identified in the presence of the full set of year dummies

to equation (13) except that now the dependent variable is the average age of the inventors on the patents granted for that firm in year t and the key right-hand-side variable is the age of the CEO, and we again control for firm fixed effects. The first column of Table 10 reports a regression of the average age of inventors on firm and year fixed effects, log employment, log sales, log patents, and CEO age, while the second column also adds dummies for inventor team size and three-digit IPC class as in the specifications in Table 9. The results, which show a positive (though only marginally significant) relationship, suggest that younger CEOs tend to hire younger inventors, indirectly corroborating the sorting effect emphasized in our theoretical model.[37] Further evidence consistent with this pattern is provided in Tables B6 and B7 in Appendix B, where we create time-varying measures of the innovation quality of new and existing inventors of the company (based on the citations of their patents in the past). We then show that both using our innovation quality measure and the tail innovation index, a younger CEO is associated with both an increase in the creativity of the innovation of continuing inventors and an increase in the quality of new inventors.

4.5 Stock of Knowledge, Opportunity Cost and Creativity of Innovations

Finally, Table 11 turns to an investigation of the additional implications of our approach highlighted in Proposition 2. We noted there that we may expect openness to disruption to be more important for companies that are technologically more advanced (as measured by the number of patents), but also that companies that have more to lose (because of the greater opportunity cost of disruption in terms of other profitable activities) may shy away from disruptive creative innovations. We investigate this issue by including the interaction between CEO age and log total patent count (as a proxy for how advanced the technology of the company is) and also the interaction between CEO age and log sales (as a proxy for company revenues that may be risked by disruptive innovations) in equation (12). According to the theoretical ideas suggested above, we expect the interaction with log total patent count to be negative, and that with sales to be positive (indicating that average manager age matters more for the creativity of innovations for companies with a significant number of patents and less for companies with high sales).

This is a demanding, as well as crude, test, since neither proxy is perfect, and moreover, log sales and log patent counts are positively correlated (the weighted correlation between the two variables is 0.7 in our sample), thus stacking the cards against finding an informative set of results.

Nevertheless, Table 11, which uses the same firm sample with annual observations as in Table 5 Panel A, provides some evidence that our theoretical expectations are borne out. In all of our specifications, the interaction between CEO age and log total patent count is negative and the interaction with log sales is positive. Moreover, these interactions are statistically significant except

[37] Interestingly, this result disappears when we do not control for firm fixed effects.

for the log patent interaction for the innovation quality measure.[38] These results thus provide some support for the hypothesis that the stock of knowledge of the company and opportunity cost effects are present and might in fact be quite important (at least quantitatively at this correlational level).

5 Conclusion

Despite a large and flourishing literature on innovation, there is relatively little work on the determinants of the creativity of innovative activity, and in particular, on innovations and patents that contribute most to knowledge. In this paper, building on Schumpeter's ideas, we suggested that openness to new ideas, disruptive innovations and unconventional practices—which we refer to as openness to disruption — may be a key determinant of creative innovations, and likewise, resistance to such disruptive behavior may hold back some of the most creative innovative activities.

We provided a simple model drawing a clear distinction between radical (more creative) innovations and incremental innovations, whereby the former combines ideas from several different lines of research and creates more significant advances (and contributions to knowledge). We showed that, because of their comparative advantage in radical innovation, younger managers tend to be employed in firms attempting radical innovations and also contribute to the creativity of innovations.

The bulk of our paper provides empirical evidence consistent with the role of openness to disruption. We use several measures to proxy for creative innovations. These include our proxy for innovation quality, which is the average number of citations per patent; two indices for creativity of innovations, which are the fraction of superstar innovators and the likelihood of a very high number of citations (in particular, fraction of the patents of a firm that are above the 99th percentile in terms of citations); and the generality index. Based on our theory, we use the age of the CEO of the company as a proxy for openness to disruption.

We find fairly consistent and robust cross-sectional and within-firm correlations between openness to disruption, proxied by CEO age, and creative innovations. They do not, however, correspond to the causal effect of CEO age on creative innovations because, as highlighted by our theoretical model, younger managers tend to be employed by firms that are more open to disruption and more creative. A simple indirect inference exercise using the structure of our model suggests that most of the empirical relationship between CEO age and creative innovations is due to these sorting effects, and the causal impact of CEO age is quite small.

Finally, our theoretical model suggests as well that the impact of openness to disruption should be larger for companies that are technologically more advanced and smaller for companies that

[38] As noted above, the main effects are evaluated at the sample mean and are typically close to the estimates reported in Table 2.

have a greater opportunity cost of disruptive innovation. The empirical patterns in our firm-level data support this prediction.

We view our paper as a first step in the study of the impact of various social and economic incentives on creative activities and, in particular, on creative innovations. Future work investigating the causal effect of manager age on creative innovations using more systematic structural estimation techniques is an obvious next step. Further study of various other firm-level or cross-country characteristics on the creativity of innovations is also a natural direction. Another fruitful direction would be to systematically investigate what types of firms and firm organizations encourage creativity and lead to more creative innovations. This would involve both theoretical and empirical analyses of the internal organization of firms and their research strategies and a study of the interplay between institutional and society-level factors and the internal organization of firms.

Appendix A: Omitted Proofs from Section 2

The Derivation of (6), (7) and (8)

A firm makes the innovation decision in each of its product lines to maximize its present discounted value, which we denote by $W_s\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)$ where $s \in \{H, L\}$, $\overrightarrow{q_f}$ is the vector of productivities of the firm, $\overrightarrow{n_f}$ is the vector of the number of incremental innovations in each of these product lines, and m_f denotes the number of product lines that firm f is operating. The value function for a low-type firm can be written as

$$rW_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) - \dot{W}_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) = \sum_{m=1}^{m_f} \left[\begin{array}{c} \max_a \left\{\pi q_{f,j_m} + \bar{q}_t f(a) - w_{a,t}\right\} \\ +\xi \left[\begin{array}{c} W_L \left(\begin{array}{c} \overrightarrow{q_f} \backslash \{q_{f,j_m}\} \cup \left\{q_{f,j_m} + \eta_{n_{f,j_m}+1}\right\}, \\ \overrightarrow{n_f} \backslash \{n_{f,j_m}\} \cup \{n_{f,j_m}+1\} \end{array} \right) \\ -W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) \end{array} \right] \\ +\tau\left[W_L\left(\overrightarrow{q_f} \backslash \{q_{f,j_m}\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\}\right) - W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)\right] \end{array} \right] + \varphi\left[W_H\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) - W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)\right]. \tag{A1}$$

The right-hand side of this value function can be explained as follows: for each product line $m = 1, ..., m_f$, the firm receives a revenue stream of $\pi q_{f,j_m}$ as a function of its productivity in this product line, q_{f,j_m}. In addition, it has a choice of the age of the manager it will hire to operate this product line (formally choosing $a \in \mathbb{R}_+ \cup \{\emptyset\}$, which is suppressed to save on notation), and if the manager's age is a, it will have additional revenue/cost savings of $\bar{q}_t f(a)$ and pay the market price for such a manager of age a at time t, $w_{a,t}$. Summing over all of its product lines gives the current revenues of the firm. In addition, the firm can undertake an innovation on the basis of the technology of each of its active product lines. Since we are looking at a low-type firm, all innovations will be incremental, thus arriving at the rate ξ. When such an innovation happens in product line m that has already undergone n_{f,j_m} incremental innovations, the mth element of $\overrightarrow{q_f}$ changes from q_{f,j_m} to $q_{f,j_m} + \eta_{n_{f,j_m}+1}$ and n goes up by one. We represent this with the arguments of the value function changing to $\overrightarrow{q_f} \backslash \{q_{f,j_m}\} \cup \left\{q_{f,j_m} + \eta_{n_{f,j_m}+1}\right\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\} \cup \{n_{f,j_m}+1\}$ (and the firm relinquishes its current value function $W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)$). The firm might also lose one of its currently active product lines to creative destruction, which happens at the endogenous rate τ, and in that case, the firm's value function changes from $W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)$ to $W_L\left(\overrightarrow{q_f} \backslash \{q_{f,j_m}\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\}\right)$ (i.e., $\overrightarrow{q_f}$ changes $\overrightarrow{q_f} \backslash \{q_{f,j_m}\}$ and $\overrightarrow{n_f}$ to $\overrightarrow{n_f} \backslash \{n_{f,j_m}\}$). Finally, the last term is due to the fact that a low-type firm switches to high-type at the flow rate φ, in which case it relinquishes its current value function and begets the value function of a high-type firm, $W_H(\overrightarrow{q_f}, \overrightarrow{n_f})$.

The value function of a high-type firm can be similarly written as

$$
\begin{aligned}
& rW_H(\overrightarrow{q_f}, \overrightarrow{n_f}) - \dot{W}_H(\overrightarrow{q_f}, \overrightarrow{n_f}) \qquad \text{(A2)} \\
= & \sum_{m=1}^{m_f} \max \left\{ \begin{array}{c} \pi q_{f,j_m} + \max_a \left\{ \bar{q}_t f(a) - w_{a,t} + \xi \left[\begin{array}{c} W_H \left(\begin{array}{c} \overrightarrow{q_f} \backslash \{q_{f,j_m}\} \cup \left\{ q_{f,j_m} + \eta_{n_{f,j_m}+1} \right\}, \\ \overrightarrow{n_f} \backslash \{n_{f,j_m}\} \cup \{n_{f,j_m}+1\} \end{array} \right) \\ -W_H(\overrightarrow{q_f}, \overrightarrow{n_f}) \end{array} \right] \right\}; \\ \pi q_{f,j_m} + \max_a \left\{ \bar{q}_t f(a) + \Lambda \theta_H \bar{q}^a \left[\begin{array}{c} \mathbb{E}W_H(\overrightarrow{q}_f \cup \{q_{j'} + \eta_0\}, \overrightarrow{n}_f \cup \{0\}) \\ -W_H(\overrightarrow{q}_f, \overrightarrow{n}_f) \end{array} \right] - w_{a,t} \right\} \end{array} \right\} \\
& + \sum_{m=1}^{m_f} \tau \left[W_H(\overrightarrow{q_f} \backslash \{q_{f,j_m}\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\}) - W_H(\overrightarrow{q_f}, \overrightarrow{n_f}) \right] \\
& + \sum_{m=1}^{m_f} \psi \Lambda \theta_H \left[\mathbb{E}W_H \left(\begin{array}{c} \overrightarrow{q}_f \cup \{q_{j'} + \eta_0\}, \\ \overrightarrow{n}_f \cup \{0\} \end{array} \right) - W_H(\overrightarrow{q}_f, \overrightarrow{n}_f) \right]
\end{aligned}
$$

The intuition for this value function is very similar to (A1) except for the possibility of a radical innovation. In particular, for each product line m, this high-type firm has a radical innovation at the flow rate $\psi \Lambda \theta_H$ regardless of its innovation strategy. In addition it has a choice between incremental and radical innovation, represented by the outer maximization. The first option here is choosing incremental innovation for product line m and is thus similar to the first line of (A1). The second option is radical innovation, and in this case the trade-off involved in the age of the manager is different, since manager age affects the arrival rate of radical innovations as shown in (5). In the case of a successful radical innovation, the value of the firm changes to $\mathbb{E}W_H\left(\overrightarrow{q_f} \cup \left\{q_{j'} + \eta_0\right\}, \overrightarrow{n_f} \cup \{0\}\right)$, where the expectation is over a product line drawn uniformly at random upon which the radical innovation will build.

Given (A1) and (A2), the value functions (6), (7) and (8) follow straightforwardly by conjecturing their forms and and verifying this conjecture.■

Proof of Proposition 1 and General Equilibrium

We first present the proof of Proposition 1, and then turn to characterizing the rest of the general equilibrium in this economy.

Proof of Proposition 1

First, substitute the equilibrium wage (9) into (7) to obtain a simplified value function for low-type firms as

$$
\begin{aligned}
rV_L(q_j, n) - \dot{V}_L(q_j, n) &= \pi q_j + \xi \left[V_L\left(q_j + \bar{q}_t \eta \alpha^{n+1}, n+1\right) - V_L(q_j, n) \right] \\
&\quad -\tau V_L(q_j, n) + \varphi \left[V_H(q_j, n) - V_L(q_j, n) \right].
\end{aligned}
$$

We next characterize the solution for this value function.

Lemma 1 *Suppose that the value function for a high-type firm takes the following form: $V_H(q_j, n) = Aq_j + \tilde{B}(n)\bar{q}_t$. Then the value function of a product line operated by a low-type firm, (7) takes the following form*

$$V_L(q_j, n) = Aq_j + B(n)\bar{q}_t \qquad \text{(A3)}$$

where

$$A \equiv \frac{\pi}{r+\tau}; \text{ and } \quad [r-g+\xi+\tau+\varphi] B(n) = \xi A \eta \alpha^{n+1} + \varphi \tilde{B}(n) + \xi B(n+1);$$

and $\tilde{B}(n)$ is defined in Lemma 2 below.

Proof of Lemma 1. We conjecture that the value function for low-type firms takes the form in (A3). Substituting this conjecture into (7), we get

$$\begin{aligned} r[Aq_j + B(n)\bar{q}_t] - B(n) g\bar{q}_t &= \pi q_j + \xi A \bar{q}_t \eta \alpha^{n+1} + \xi [B(n+1)\bar{q}_t - B(n)\bar{q}_t] \\ &\quad -\tau A q_j - \tau B(n)\bar{q}_t + \varphi \left[A q_j + \bar{q}_t \tilde{B}(n) - A q_j - B(n)\bar{q}_t \right]. \end{aligned}$$

Equating the coefficients on q_j and $\bar{q}_t$, we obtain

$$rAq_j = \pi q_j - \tau A q_j; \text{ and}$$

$$rB(n)\bar{q}_t - B(n) g\bar{q}_t = \xi A \bar{q}_t \eta \alpha^{n+1} + \bar{q}_t \xi [B(n+1) - B(n)] - \tau B(n)\bar{q}_t + \bar{q}_t \varphi \left[\tilde{B}(n) - B(n) \right].$$

Solving these equations for A and $B(n)$, taking $\tilde{B}(n)$ as given (it will be determined in Lemma 2) completes the proof. ■

The form of the value function in (A3) is intuitive. It depends linearly on current productivity, q_j, which determines the current flow of profits. It also depends on current economy-wide technology, $\bar{q}_t$, since all innovations, including incremental ones, build on this. Finally, it is decreasing in n (because $B(n)$ is decreasing as we will see) since a higher n implies that the next incremental innovation will increase productivity by less.

Next, substitute (10) into (8) to obtain a simplified form of the value function of a product line operated by a high-type firm as

$$\begin{aligned} rV_H(q_j, n) - \dot{V}_H(q_j, n) &= \max \left\{ \begin{array}{c} \pi q_j + \xi \left[V_H\left(q_j + \bar{q}_t \eta \alpha^{n+1}, n+1\right) - V_H(q_j, n) \right]; \\ \pi q_j + \Lambda \theta_H \bar{q}^{a^*} \mathbb{E} V_H(\bar{q}_t) \end{array} \right\} \\ &\quad -\tau V_H(q_j, n) + \psi \Lambda \theta_H \mathbb{E} V_H(\bar{q}_t). \end{aligned}$$

We next characterize the solution to this value function and also determine the allocation of managers to different product lines (and to incremental and radical innovations).

Lemma 2 *The value function in (8) takes the following form*

$$V_H(q_j, n) = Aq_j + \bar{q}_t \tilde{B}(n), \tag{A4}$$

where A and $B(n)$ are as defined in Proposition 1) and $\tilde{B}(n)$ is given by

$$[r-g+\tau] \tilde{B}(n) = \psi \left[A(1+\eta) + \tilde{B}(0) \right] + \left\{ \begin{array}{ll} \xi \left[\tilde{A} \eta \alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n) \right] & \text{for } n < n^* \\ \Lambda \theta_H \bar{q}^{a^*} \left[(1+\eta) \tilde{A} + \tilde{B}(0) \right] & \text{for } n \geq n^* \end{array} \right., \tag{A5}$$

where

$$n^* = \min_{n' \in \mathbb{Z}_+} n' \text{ such that } \xi \left[A \eta \alpha^{n'+1} + \tilde{B}(n'+1) - \tilde{B}(n') \right] \leq \Lambda \theta_H \bar{q}^{a^*} \left[(1+\eta) A + \tilde{B}(0) \right]. \tag{A6}$$

Proof of Lemma 2. We now conjecture that the value function for high-type firms takes the form in (A4), and substitute this into (8) to obtain

$$(r+\tau)\left[Aq_j+\bar{q}_t\tilde{B}(n)\right]-g\bar{q}_t\tilde{B}(n)=\pi q_j+\psi\Lambda\theta_H\left[A\bar{q}_t+A\eta\bar{q}_t+\bar{q}_t\tilde{B}(0)\right]+\max\left\{\begin{array}{c}\bar{q}_t\xi\left[A\eta\alpha^{n+1}+\tilde{B}(n+1)-\tilde{B}(n)\right];\\ \Lambda\theta_H\bar{q}^{a^*}\left[A\bar{q}_t+A\eta\bar{q}_t+\bar{q}_t\tilde{B}(0)\right]\end{array}\right\}.$$

Once again equating coefficients, we obtain $A=\frac{\pi}{r+\tau}$ and

$$(r-g+\tau)\tilde{B}(n)=\psi\Lambda\theta_H\left[A(1+\eta)+\tilde{B}(0)\right]+\max\left\{\begin{array}{c}\xi\left[A\eta\alpha^{n+1}+\tilde{B}(n+1)-\tilde{B}(n)\right];\\ \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]\end{array}\right\}. \tag{A7}$$

Let us next define $\hat{B}(n)$ as the solution to the equation

$$(r-g+\tau)\hat{B}(n)=\psi\Lambda\theta_H\left[A(1+\eta)+\tilde{B}(0)\right]+\xi\left[A\eta\alpha^{n+1}+\hat{B}(n+1)-\hat{B}(n)\right].$$

Under the hypothetical scenario where the max operator in (A7) always picks the first term, we have $\tilde{B}(n)=\hat{B}(n)$. Collecting terms,

$$\hat{B}(n)=\beta\hat{\psi}/\xi+\beta A\eta\alpha^{n+1}+\beta\hat{B}(n+1), \tag{A8}$$

where $\beta=\frac{\xi}{(r-g+\tau+\xi)}$ and $\hat{\psi}=\psi\Lambda\theta_H\left[A(1+\eta)+\tilde{B}(0)\right]$. Note that (A8) defines a contraction (in particular, it satisfies the monotonicity and discounting sufficient conditions of Blackwell, e.g., Theorem 3.3 in Stokey and Lucas, 1989). Since, in addition, $\beta A\eta\alpha^{n+1}$ is strictly decreasing in n, $\hat{B}(n)$ is strictly decreasing as well (e.g., Theorem 4.7 in Stokey and Lucas, 1989). Now if $n^*=\infty$ (meaning that incremental innovations were always optimal), then we would have $\tilde{B}(n)=\hat{B}(n)$.

The other option in the max operator, $\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]$, does not depend on n and is strictly positive. Moreover, note that for n large, $\hat{B}(n)$ limits to $\frac{\hat{\psi}}{r-g+\tau}$. This is strictly less than what a firm can obtain by switching to radical innovation at n, $\tilde{B}(n)=\frac{\hat{\psi}+\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]}{r-g+\tau}$. Therefore, there exists a smallest integer n^* (which could be zero) such that $\xi\left[A\eta\alpha^{n+1}+\tilde{B}(n+1)-\tilde{B}(n)\right]>\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]$ for all $n<n^*$, which verifies the definition of n^* in (A6). By the definition of n^*, we have that $\xi\left[A\eta\alpha^{n^*+1}+\tilde{B}(n^*+1)-\tilde{B}(n^*)\right]\leq\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]$. Now there are two cases to consider:

1. $\xi\left[A\eta\alpha^{n^*+1}+\tilde{B}(n^*+1)-\tilde{B}(n^*)\right]<\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]$. Then, at n^* it is strictly optimal to switch to radical innovation, and thus $\lambda_{n^*}=1$.

2. $\xi\left[A\eta\alpha^{n^*+1}+\tilde{B}(n^*+1)-\tilde{B}(n^*)\right]=\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]$. In this case, firms are indifferent between incremental and radical innovation at n^*, and thus $\lambda_{n^*}\in[0,1]$.

We summarize these two cases with the complementary slackness condition

$$\begin{array}{c}\lambda_{n^*}\leq 1,\ \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]-\xi\left[A\eta\alpha^{n^*+1}+\tilde{B}(n^*+1)-\tilde{B}(n^*)\right]\geq 0 \text{ and}\\ \left[\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]-\xi\left[A\eta\alpha^{n^*+1}+\tilde{B}(n^*+1)-\tilde{B}(n^*)\right]\right]\times(1-\lambda_{n^*})=0.\end{array} \tag{A9}$$

Observe also that with the same argument we used for $\hat{B}(n)$, $\tilde{B}(n)$ can be proved to be (weakly) decreasing in n as claimed following the proof of Lemma 2 (since $\tilde{B}(n)$ is defined by a contraction which maps decreasing functions into themselves).

Finally, we prove that $\lambda_n = 1$ for all $n > n^*$. Suppose not. To obtain a contradiction, define $\tilde{b}(n) = A\eta\alpha^n + \tilde{B}(n) - \tilde{B}(n-1)$, and subtract (A7) lagged once from itself, which gives

$$(r-g+\tau)\,\tilde{b}(n)+\max\left\{\begin{array}{c}\xi\tilde{b}(n);\\ \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]\end{array}\right\} = (r-g+\tau)\,A\eta\alpha^n+\max\left\{\begin{array}{c}\xi\tilde{b}(n+1);\\ \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]\end{array}\right\}. \tag{A10}$$

We next verify that $\tilde{b}(n)$ is a contraction over the set of continuous functions that are decreasing in n. To show that it is a contraction, we just verify the sufficiency conditions of Blackwell (e.g., Theorem 3.3 in Stokey and Lucas, 1989). Let the left-hand side of (A10) be denoted by $F(\tilde{b}(n))$. Clearly, F is strictly monotone, and thus has a strictly monotone inverse F^{-1}. Then, (A10) can be written as

$$\tilde{b}(n) = F^{-1}\left[(r-g+\tau)\,A\eta\alpha^n + \max\left\{\xi\tilde{b}(n+1);\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]\right\}\right], \tag{A11}$$

and thus

$$\tilde{b}(\cdot) = T(\tilde{b}(\cdot)),$$

where the operator $T : D(\mathbb{Z}_+) \to D(\mathbb{Z}_+)$ is defined by the right-hand side of the previous expression and $D(\mathbb{Z}_+)$ is the set of decreasing continuous functions over $\mathbb{Z}_+$. Since $(r-g+\tau)\,A\eta\alpha^n$ is strictly decreasing and F^{-1} is increasing, T maps decreasing continuous functions into themselves (and in fact, it maps them into strictly decreasing functions). That T satisfies monotonicity is immediate.

To see that it satisfies the discounting condition as well, we will show that for any $c > 0$, $T(\tilde{b}(\cdot)+c) \leq T(\tilde{b}(\cdot)) + \beta c$ for some $\beta < 1$. First suppose that $\xi\tilde{b}(n+1) + \xi c \leq \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]$ (case (i)). In this case,

$$\begin{aligned} T(\tilde{b}(n+1)+c) &= F^{-1}\left[(r-g+\tau)\,A\eta\alpha^n + \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]\right] \\ &= T(\tilde{b}(n+1)). \end{aligned} \tag{A12}$$

(Note that here we are using $T(\tilde{b}(n+1))$ to designate the value of $\tilde{b}(n)$ as given by (A11) evaluated at a specific n, while $T(\tilde{b}(\cdot))$ denotes the entire mapping).

Suppose, alternatively, that we are in case (ii), where $\xi\tilde{b}(n+1) + \xi c > \Lambda\theta_H A\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]$. In this case, since $\tilde{b}(n)$ is decreasing, the fact that $\xi\tilde{b}(n+1)+\xi c > \Lambda\theta_H A\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]$ implies that $\xi\tilde{b}(n)+\xi c > \Lambda\theta_H A\bar{q}^{a^*}\left[(1+\eta)\,A+\tilde{B}(0)\right]$, and thus

$$\begin{aligned} T(\tilde{b}(n)+c) &= F^{-1}[(r-g+\tau)\,A\eta\alpha^n + \xi\tilde{b}(n+1) + \xi c] \\ &= \frac{r-g+\tau}{r-g+\tau+\xi}A\eta\alpha^n + \frac{\xi}{r-g+\tau+\xi}\tilde{b}(n+1) + \frac{\xi}{r-g+\tau+\xi}c \\ &\leq T(\tilde{b}(n+1)) + \frac{\xi}{r-g+\tau+\xi}c. \end{aligned} \tag{A13}$$

where the second line follows from the observation just proceeding the equation, since in this case,

$F^{-1}(b) = b/(r - g + \tau + \xi)$, and the third line follows because

$$\begin{aligned} T(\tilde{b}(n+1)) &= F^{-1}\left[(r-g+\tau)\, A\eta\alpha^n + \max\left\{\xi\tilde{b}(n+1); \Lambda\theta_H \bar{q}^{a^*}\left[(1+\eta)\, A + \tilde{B}\,(0)\right]\right\}\right] \\ &\geq F^{-1}\left[(r-g+\tau)\, A\eta\alpha^n + \xi\tilde{b}(n+1)\right]. \end{aligned}$$

Therefore, combining cases (i) and (ii), i.e. (A12) and (A13), we have that

$$T(\tilde{b}(\cdot) + c) \leq T(\tilde{b}(\cdot)) + \frac{\xi}{r - g + \tau + \xi} c.$$

Then, by setting $\beta = \frac{\xi}{r-g+\tau+\xi} < 1$, the discounting condition follows.

We thus conclude that $\tilde{b}$ is a decreasing continuous function. Moreover, because T maps decreasing functions into strictly decreasing functions, $\tilde{b}$ is in fact strictly decreasing. Next, by the definition of n^*, $\tilde{b}(n^*) \leq \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\, A + \tilde{B}\,(0)\right]$, and since $\tilde{b}$ is strictly decreasing, we have $\tilde{b}(n^*+1) < \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\, A + \tilde{B}\,(0)\right]$. But this yields a contradiction with $\lambda_{n^*+1} < 1$, establishing the desired result. ■

The intuition for this high-type value function is similar to that for Proposition 1, except that the dependence on the number of prior innovations in the current technology cluster, n, is more complicated since when n exceeds n^*, a high-type firm will switch to radical innovation. This critical value n^* is given by (A6) as the smallest integer such that pursuing incremental innovations is no longer strictly optimal.

Proposition 1 then follows directly from Lemma 1 and 2.■

General Equilibrium and the Stationary Equilibrium

Consumer maximization yields the usual Euler equation, $\frac{\dot{C}(t)}{C(t)} = \frac{1}{\nu}(r(t) - \rho)$,where g is the growth rate of the economy, as well as a standard transversality condition, $\lim_{t\to\infty} A(t)e^{-\int_0^t r(s)ds} = 0$, where $A(t)$ is the value of total assets of households at time t, which come from their ownership of the firms in the economy. In a stationary equilibrium, we have $r(t) = r$ and $\dot{C}(t)/C(t) = g$ (because, from (2), $K(t)$ grows at the same rate as $Y(t)$, g). This implies

$$r = \rho + \nu g. \tag{A14}$$

Moreover, since $A(t)$ will grow at the rate g, the transversality condition simply requires $r > g$, or

$$(1 - \nu)g < \rho, \tag{A15}$$

which we impose as an assumption.

In a stationary equilibrium, the threshold, n^*, is constant. This is because the value function increases linearly in $\bar{q}_t$, but the knowledge stock and wages of managers also increase linearly, and in the stationary equilibrium, these two forces balance out, ensuring that n^* is constant while V_H increases linearly in $\bar{q}_t$.

Given the form of V_H, $\mathbb{E}V_H(\bar{q}_t)$, the value of a new radical innovation, can be written as

$$\begin{aligned} \mathbb{E}V_H(\bar{q}_t) &= \mathbb{E}\left[\tilde{A}q_j + \tilde{A}\eta\bar{q}_t + \bar{q}_t\tilde{B}\,(0)\right] = [\tilde{A}\,(1+\eta) + \tilde{B}\,(0)]\bar{q}_t \\ &\equiv v\bar{q}_t, \end{aligned}$$

where the last line defines v. Then the equilibrium wage schedule simplifies to

$$w_{a,t} = \begin{cases} f(a)\,\bar{q}_t & \text{for} \quad a > a^* \\ [f(a) + \Lambda\theta_H(\bar{q}^a - \bar{q}^{a^*})v]\bar{q}_t & \text{for} \quad a \leq a^* \end{cases} . \tag{A16}$$

and is thus also linear in $\bar{q}_t$.

We next characterize the stationary distribution of product lines in this economy in terms of the types of their owners and also in terms of the prior number of incremental innovations, and then use these distributions to determine the aggregate growth rate of the economy in the stationary equilibrium. Let us denote the fraction of product lines occupied by s-type firms (for $s \in \{L, H\}$) with n prior incremental innovations by μ_n^s (these are not functions of time as we are focusing on a stationary equilibrium). Denote the total creative destruction from s-type firms by τ^s. The stationary distribution of product lines is determined by standard flow equations equating inflows and outflows from each state. For high types, these take the form

$$\begin{array}{rcll} \underline{\textsc{Outflow}} & & \underline{\textsc{Inflow}} & \\ \left(\tau^L + \xi\right)\mu_0^H & = & \tau^H\left(1 - \mu_0^H\right) + \varphi\mu_0^L & \text{for} \quad n = 0 \\ \left(\tau^L + \tau^H + \xi(1-\lambda_n)\right)\mu_n^H & = & \xi\mu_{n-1}^H(1-\lambda_{n-1}) + \varphi\mu_n^L & \text{for} \quad n > 0 \end{array} . \tag{A17}$$

Consider the first line corresponding to $n = 0$. Outflows from this state (which refers to products with $n = 0$ operated by high-type firms) come from two sources. First, there is creative destruction coming from low-type firms (entrants), which takes place at the rate τ^L per product line (and hence multiplied by μ_0^H). Second, the high-type firm operating this product line has a successful incremental innovation, which takes place at the rate ξ (similarly multiplied by μ_0^H). Inflows into this state are due to creative destruction coming from high-type firms, which takes place at the rate τ^H (multiplied by the fraction of all product lines except those that are already in this state, thus $\left(1 - \mu_0^H\right)$), or due to a low-type firm operating a product line at $n = 0$ changing its type to high, which adds the flow rate $\varphi\mu_0^L$. The other lines are explained similarly, taking into account the fact that at n, firms pursue incremental innovations with probability $1 - \lambda_n$, and that inflows into product line n will depend on the probability with which high-type firms in product line $n - 1$ pursue incremental innovations, given by $1 - \lambda_{n-1}$.[39]

The flow equations for the low-type product lines can be written similarly,

$$\begin{array}{rcll} \underline{\textsc{Outflow}} & & \underline{\textsc{Inflow}} & \\ \left(\tau^H + \xi + \varphi\right)\mu_0^L & = & (1 - \mu_0^L)\tau^L & \text{for} \quad n = 0 \\ \left(\tau^L + \tau^H + \xi + \varphi\right)\mu_n^L & = & \xi\mu_{n-1}^L & \text{for} \quad n > 0 \end{array} .$$

Since these flow equations only depend on τ^H and τ^L, they can be solved out explicitly, which is particularly convenient for our proof of existence of a stationary equilibrium. Namely, we have

$$\mu_n^L = \left(\frac{\xi}{\tau + \xi + \varphi}\right)^{n+1} \frac{\tau^L}{\xi}. \tag{A18}$$

The creative destruction rates from low-type and high-type firms, in turn, can be computed as

$$\tau^L = x\,(1 - \zeta) \quad \text{and} \quad \tau^H = x\zeta + M\int_0^{a^*} \Lambda\theta_H \bar{q}^a dF(a) + \psi\Lambda\theta_H \textstyle\sum_{n=0}^{\infty} \mu_n^H, \tag{A19}$$

[39] These equations are written under the assumption that $n^* > 0$, and can be adjusted straightforwardly for the case in which $n^* = 0$.

where x is the (exogenous) entry rate, $F(a)$ denotes the stationary distribution of manager age, a^* is the threshold below which managers are hired by firms to perform radical innovations, and $\psi\Lambda\theta_H \sum_{n=0}^{\infty} \mu_n^H$ is the rate of radical innovations for high-type firms which applies regardless of whether or not they pursue a radical innovation strategy. Low-type firms, on the other hand, generate creative destruction only when they initially enter the economy (since they do not engage in radical innovation). Given these quantities, the total creative destruction rate of the economy is given as

$$\tau = \tau^L + \tau^H. \tag{A20}$$

Notice that we wrote the creative destruction rate coming from high-type firms as a function of the age threshold for young managers, a^*. In equilibrium, the measure of young managers working on radical innovations must equal the measure of firms attempting radical innovations, which gives us the additional equilibrium condition

$$MF(a^*) = \sum\nolimits_{n=0}^{\infty} \mu_n^H (1 - \lambda_n). \tag{A21}$$

The final equation we need is for the aggregate growth rate. Let us combine (1) with (2) to obtain

$$Y = \frac{L}{1-\beta} \left[\frac{(1-\beta)}{\gamma} \right]^{\frac{1-\beta}{\beta}} \bar{q}.$$

The growth rate of the economy is then equal to the growth of the average quality $\bar{q}_t$. After a time interval $\Delta t > 0$, the average quality is given by

$$\bar{q}_{t+\Delta t} = \bar{q}_t + \eta \bar{q}_t \tau \Delta t + \bar{q}_t \eta \xi \Delta t \left[\sum\nolimits_{n=0}^{\infty} \mu_n^H (1 - \lambda_n) \alpha^{n+1} + \sum\nolimits_{n=0}^{\infty} \mu_n^L \alpha^{n+1} \right] + o(\Delta t),$$

where we have used the fact that all radical innovations come from creative destruction, which takes place at the rate τ, and $o(\Delta t)$ denotes terms that are second order in Δt. The growth rate of the economy in the stationary equilibrium can then be computed as

$$g = \eta \tau + \eta \xi \left[\sum\nolimits_{n=0}^{\infty} \mu_n^H (1 - \lambda_n) \alpha^{n+1} + \sum\nolimits_{n=0}^{\infty} \mu_n^L \alpha^{n+1} \right]. \tag{A22}$$

Finally, we establish the existence of a stationary equilibrium. To do this, first let us suppress the explicit determination of n^*, and instead introduce the equivalent of the complementary slackness condition, (A9), which determined λ_{n^*}, for each n,

$$\begin{aligned}
\lambda_n = 0 \text{ if } \Lambda\theta_H \bar{q}^{a^*} \left[(1+\eta) A + \tilde{B}(0) \right] - \xi \left[A\eta\alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n) \right] < {} & 0, \qquad \text{(A23)} \\
\lambda_n = 1 \text{ if } \Lambda\theta_H \bar{q}^{a^*} \left[(1+\eta) A + \tilde{B}(0) \right] - \xi \left[A\eta\alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n) \right] > {} & 0, \text{ and} \\
\left[\Lambda\theta_H \bar{q}^{a^*} \left[(1+\eta) A + \tilde{B}(0) \right] - \xi \left[A\eta\alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n) \right] \right] \times (1 - \lambda_{n^*}) = {} & 0.
\end{aligned}$$

Let us also suppress μ_n^L as endogenous variables, instead using the explicit solutions in (A18), and eliminate the stationary equilibrium value of the interest rate, r, from the equilibrium condition (A14). This leaves us with the endogenous variables to be determined as: $\{\mu_n^H, \lambda_n\}_{n=0}^{\infty}$, τ^L, τ^H, a^* and g, which satisfy equations (A17), (A19), (A21), (A22) and (A23). Note further that there is an upper bound on the creative destruction rate, $\bar{\tau}$, which is obtained by (counterfactually) assuming that all product lines successfully pursue radical innovation: $\bar{\tau} = (1+\psi) \Lambda\theta_H + x$. Similarly, there is also an upper bound on the growth rate of the economy, corresponding to the growth rate that would follow when the creative destruction rate takes its upper bound value and all innovations have the maximal step size, η; this gives the upper bound as $\bar{g} \equiv (\bar{\tau} + \xi)\,\eta$. We can then take the growth rate and the creative destruction rate of the economy to lie in the compact sets, $[0, \bar{g}]$ and $[0, \bar{\tau}]$, respectively.

We can further reduce the system from an infinite-dimensional one to a finite-dimensional one by noting that $\tilde{b}(n)$ is decreasing in $\tau - g$, thus by taking $g = \bar{g}$ and $\tau = 0$, we can compute an upper bound for the threshold n^* as $\bar{n}$. Given this upper bound, we derive an upper bound for a^* given by $\bar{a}$ from (A21) by truncating the summation on the right-hand side at $\bar{n}$, and setting $\lambda_n = 0$ and $\mu_n^H = 1$ for all $n \leq \bar{n}$. Thus we have a finite-dimensional mapping Φ : $[0,1]^{\bar{n}} \times [0,\bar{\tau}]^2 \times [0,\bar{a}] \times [0,\bar{g}] \rightarrow [0,1]^{\bar{n}} \times [0,\bar{\tau}]^2 \times [0,\bar{a}] \times [0,\bar{g}]$, which maps a compact set into itself. Moreover, (A17), (A19), (A21), and (A22) are continuous functions, while (A23) defines an upper hemi-continuous, closed- and convex-valued correspondence. Therefore, Φ is an upper hemi-continuous, closed- and convex-valued correspondence, and by Kakutani's fixed point theorem, it has a fixed point. By construction, this fixed point is a stationary equilibrium. ■

Proof of Proposition 2

The value of a product line operated by low- and high-type firms can now be written, respectively, as:

$$\begin{aligned} rV_L(q_j,n) - \dot{V}_L(q_j,n) &= \max_a \{\pi q_j + \bar{q}_t f(a) - w_{a,t}\} + \xi \left[V_L\left(q_j + \eta_{n+1}, n+1\right) - V_L(q_j,n)\right] \\ &\quad -\tau V_L(q_j,n) + \varphi\left[V_H(q_j,n) - V_L(q_j,n)\right], \\ &\quad \text{and} \\ rV_H(q_j,n) - \dot{V}_H(q_j,n) &= \max\left\{ \begin{array}{c} \pi q_j + \max_a \left\{ \bar{q}_t f(a) - w_{a,t} + \xi \left[\begin{array}{c} V_H\left(q_j + \eta_{n+1}, n+1\right) \\ -V_H(q_j,n) \end{array} \right] \right\}; \\ \pi q_j + \max_{a \geq 0} \{\bar{q}_t f(a) + \Lambda \theta_H \bar{q}^a \mathbb{E} V_H(t) - w_{a,t}\} \end{array} \right\} \\ &\quad -\tau V_H(q_j,n) + \psi \Lambda \theta_H \mathbb{E} V_H(t). \end{aligned}$$

Here note that, with a slight abuse of notation, we wrote $\mathbb{E}V_H(t)$ instead of $\mathbb{E}V_H(\bar{q}_t)$ for the value of a new radical innovation, since this depends in general not just on average current productivity in the economy, $\bar{q}_t$, but also on the distribution of product lines across different states. All the same, in the stationary equilibrium it will clearly grow at the same rate as $\bar{q}_t$, g. Second, η_n is now a function of both the current productivity of the firm and the average current productivity in the economy, $\bar{q}_t$.

With an argument similar to that in the previous subsection, the equilibrium wage schedule for managers will be given by

$$w_{a,t} = \begin{cases} f(a)\,\bar{q}_t & \text{for} \quad a > a^* \\ f(a)\,\bar{q}_t + \Lambda \theta_H \left[\bar{q}^a - \bar{q}^{a^*}\right] \mathbb{E} V_H(t) & \text{for} \quad a \leq a^* \end{cases}$$

This enables us to write simplified versions of the value functions as:

$$\begin{aligned} rV_L(q_j,n) - \dot{V}_L(q_j,n) &= \pi q_j + \xi\left[V_L\left(q_j + \eta_{n+1}, n+1\right) - V_L(q_j,n)\right] \\ &\quad -\tau V_L(q_j,n) + \varphi\left[V_H(q_j,n) - V_L(q_j,n)\right] \end{aligned}$$

and

$$
\begin{aligned}
rV_H(q_j,n) - \dot{V}_H(q_j,n) &= \max\left\{ \begin{array}{c} \pi q_j + \xi\left[V_H(q_j+\eta_{n+1}, n+1) - V_H(q_j,n)\right]; \\ \pi q_j + \Lambda\theta_H \bar{q}^{a^*}\mathbb{E}V_H(t) \end{array} \right\} \\
&\quad -\tau V_H(q_j,n) + \psi\Lambda\theta_H \mathbb{E}V_H(t).
\end{aligned}
$$

Writing explicitly $\eta_{n+1,t}(q_{n,t})$ as the incremental improvement in productivity starting from quality $q_{n,t}$ that has been improved n times already and average quality in the economy is $\bar{q}_t$ (subsumed in the time argument t), we have

$$
\begin{aligned}
(r+\tau)V_H(q_{n,t},n) - \dot{V}_H(q_{n,t},n) &= \pi q_{n,t} + \max\left\{ \begin{array}{c} \xi\left[V_H(q_{n,t}+\eta_{n+1,t}(q_{n,t}), n+1) - V_H(q_{n,t},n)\right]; \\ \Lambda\theta_H \bar{q}^{a}\mathbb{E}V_H(t) \end{array} \right\} \\
&\quad +\psi\Lambda\theta_H \mathbb{E}V_H(t).
\end{aligned}
$$

The threshold number of incremental innovations as a function of current productivity, $n_t^*(q)$ equivalently defines a threshold value of productivity $q_{n,t}^*$ as a function of the number of incremental innovations. Clearly this threshold productivity level is defined as the value that sets the two terms in the max operator equal to each other. Thus

$$
V_H\left(q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*), n+1\right) - V_H\left(q_{n,t}^*, n\right) = \frac{\Lambda\theta_H\bar{q}^a}{\xi}\mathbb{E}V_H(t), \tag{A24}
$$

and at this value, we also have

$$
(r+\tau)V_H\left(q_{n,t}^*, n\right) - \dot{V}_H\left(q_{n,t}^*, n\right) = \pi q_{n,t}^* + \Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t) + \psi\Lambda\theta_H\mathbb{E}V_H(t). \tag{A25}
$$

Now we will consider two alternative cases:

Case 1:

$$
q_{n+1,t}^* \geq q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*). \tag{A26}
$$

This condition implies that if a particular high-type firm finds it optimal to switch to radical innovation today, but instead undertakes a successful incremental innovation (as a deviation off-the-equilibrium path), then subsequently it will still want to immediately switch to radical innovation.

Under this case, we have

$$
\begin{aligned}
& (r+\tau)V_H\left(q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*), n+1\right) - \dot{V}_H\left(q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*), n+1\right) \\
= \; & \pi q_{n,t}^* + \pi\eta_{n+1,t}(q_{n,t}^*) + \Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t) + \psi\Lambda\theta_H\mathbb{E}V_H(t).
\end{aligned} \tag{A27}
$$

This follows from the fact that, by definition, in this case, at $q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*)$, the firm will want to switch to radical innovation.

Now differentiating (A24) with respect to time, we have

$$
\dot{V}_H\left(q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*), n+1\right) - \dot{V}_H\left(q_{n,t}^*, n\right) = \frac{\Lambda\theta_H\bar{q}^a}{\xi}\partial\mathbb{E}V_H(t)/\partial t = \frac{\Lambda\theta_H\bar{q}^a}{\xi}g\mathbb{E}V_H(t), \tag{A28}
$$

where, in the second line, we have used the fact that in a stationary equilibrium $\mathbb{E}V_H(t)$ grows at the rate g. Subtracting (A25) from (A27) and using (A28), we obtain:

$$
(r+\tau)[V_H\left(q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*), n+1\right) - V_H\left(q_{n,t}^*, n\right)] = \pi\eta_{n+1,t}(q_{n,t}^*) + \frac{\Lambda\theta_H\bar{q}^a}{\xi}g\mathbb{E}V_H(t). \tag{A29}
$$

Then, combining (A24) and (A29) we can derive

$$\pi\eta_{n+1,t}(q^*_{n,t}) = \frac{r-g+\tau}{\xi}\Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t). \tag{A30}$$

In this case, for all q less than $q^*_{n,t}$, it is optimal to switch to radical innovation.

Now let us define

$$v_t \equiv \frac{r-g+\tau}{\pi\xi}\Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t), \tag{A31}$$

which is independent of both q and n. Using (A31) equation (A30) can be written as

$$[\kappa\bar{q}_t + (1-\kappa)q^*_{n,t}]\eta\alpha^{n+1} = v_t, \tag{A32}$$

or

$$q^*_{n,t} = \frac{v_t/\eta\alpha^{n+1} - \kappa\bar{q}_t}{1-\kappa}. \tag{A33}$$

This equation implies that $q^*_{n,t}$ is increasing in n or equivalently that $n^*_t(q)$ is increasing in q.

We next derive the condition under which (A26) indeed applies. For this reason, note that from (A32) written for $n+2$ incremental innovations, we have

$$q^*_{n+1,t} = \frac{v_t/\eta\alpha^{n+2} - \kappa\bar{q}_t}{1-\kappa}. \tag{A34}$$

Combining equations (A33) and (A34), we obtain that (A26) is satisfied if

$$(1-\kappa)\eta\alpha^{n+2} + \alpha \leq 1. \tag{A35}$$

Thus whenever (A35) holds (and we are in Case 1), we have the desired result that $n^*_t(q)$ is increasing in q. We next establish that whenever the converse of (A35) holds, the same result applies.

Case 2:

$$q^*_{n+1,t} - \eta_{n+1,t}(q^*_{n,t}) < q^*_{n,t}. \tag{A36}$$

This implies that if a high-type firm is indifferent between radical and incremental innovation at $n+1^{st}$ prior incremental innovations at time t, then it would have preferred to switch to radical innovation at n^{th} prior incremental innovations. This condition is clearly the complement of (A26).

In this case, start with $q^*_{n+1,t}$, which satisfies (A27). Under condition (A36), $q^*_{n,t}$ satisfies (A25), so we again arrive at (A24), (A30) and (A33). But then from (A33) $q^*_{n,t}$ is increasing in n or $n^*_t(q)$ is increasing in q.

We next verify that Case 2 applies for the complement of the parameter values for which (A35) holds. Note that the same expressions for $q^*_{n+1,t}$ as in (A34) again applies under Case 2. Thus the condition for (A36) to be satisfied, with an identical argument, is

$$(1-\kappa)\eta\alpha^{n+2} + \alpha > 1,$$

which is indeed the complement of (A35).

Consequently, regardless of whether (A35) or its converse holds, equation (A33) applies, and $q^*_{n,t}$ is increasing in n (or equivalently, $n^*_t(q)$ is increasing in q). This completes the proof.■

References

Abrams, D. S., U. Akcigit, and J. Popadak (2013). Patent Value and Citations: Creative Destruction or Defensive Disruption? National Bureau of Economic Research Working Paper 19763.

Acemoglu, D., U. Akcigit, N. Bloom, and W. R. Kerr (2013). Innovation, Reallocation and Growth. National Bureau of Economic Research Working Paper 18993.

Akcigit, U. and W. R. Kerr (2010). Growth through Heterogeneous Innovations. National Bureau of Economic Research Working Paper 16443.

Akcigit, U. and W. R. Kerr (2015). Growth through Heterogeneous Innovations, Second Version. Penn Institute for Economic Research Working Paper 15-020.

Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58(2), 277-297.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. *NBER Chapters*, 609-626.

Azoulay, P., G. Manso and J. G. Zivin (2011). Incentives and Creativity: Evidence from the Academic Life Sciences. *RAND Journal of Economics* 42(3), 527-544.

Azoulay, P., J. G. Zivin, and J. Wang (2010). Superstar Extinction. *Quarterly Journal of Economics*. 125(2), 549-589.

Barker, V. L., and G. C. Mueller (2002). CEO Characteristics and Firm R&D Spending. *Management Science,* 48(6), 782-801.

Benmelech, E., and C. Frydman (2014). Military CEOs. National Bureau of Economic Research Working Paper 19782.

Bennedsen, M., F. Pérez-González, and D. Wolfenzon (2008). Do CEOs Matter? Stanford GSB Working Paper.

Bertrand, M. and A. Schoar (2003). Managing with Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics,* 118(4), 1169-1208.

Bloom, N., and J. Van Reenen (2007). Measuring and Explaining Management Practices across Firms and Countries. *Quarterly Journal of Economics,* 122(4), 1351-1408.

Bloom, N., and J. Van Reenen (2010). Why Do Management Practices Differ across Firms and Countries? *Journal of Economic Perspectives,* 24(1), 203-224.

Chari, V.V., and H. Hopenhayn (1991). Vintage Human Capital, Growth, and the Diffusion of New Technology. *Journal of Political Economy*, 99(6), 1142–1165.

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

Fogli A. and L. Veldkamp (2013). Germs, Social Networks and Growth. National Bureau of Economic Research Working Paper 18470.

Foster, L., J. Haltiwanger, and C. J. Krizan (2001). Aggregate Productivity Growth: Lessons from Microeconomic Evidence. *New Directions in Productivity Analysis* (eds. E. Dean, M. Harper, and C. Hulten), University of Chicago Press, 303-372.

Galenson D. and B. Weinberg (1999). Age and the Quality of Work: The Case of Modern American Painters. National Bureau of Economic Research Working Paper 7122.

Galenson D. and B. Weinberg (2001). Creating Modern Art: The Changing Careers of Painters in France From Impressionism to Cubism. *American Economic Review*, 91 (4), 1063-1071.

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy,* 108(5), 874-904.

Garicano, L., and E. Rossi-Hansberg (2004). Inequality and the Organization of Knowledge. *American Economic Review,* 94(2), 197-202.

Gorodnichenko Y. and G. Roland (2012). Culture, Institutions and the Wealth of Nations. UC Berkeley Working Paper.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. National Bureau of Economic Research Working Paper 8498.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999). Citation Frequency and the Value of Patented Inventions. *Review of Economics and Statistics*, 81(3), 511-515.

Hofstede, G. H. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations.* Sage.

Hurst, E. and B. Pugsley (2011). What Do Small Businesses Do? *Brookings Papers on Economic Activity*, Fall 2011.

Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *Review of Economic Studies,* 76 (1), 283-317.

Jones, B. F. (2010). Age and Great Invention. *Review of Economics and Statistics,* 92 (1), 1-14.

Jones, B. F. and B. Weinberg (2011). A Dynamics and Scientific Creativity. *Proceedings of the National Academy of Sciences*, 108 (47), 18910-18914.

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

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2012). Technological Innovation, Resource Allocation, and Growth. National Bureau of Economic Research Working Paper 17769.

Lai, R., A. D'Amour, A. Yu, Y. Sun, and L. Fleming (2009). Disambiguation and Co-author Networks of the US Patent Inventor Database. Harvard Business School Working Paper.

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

Lucas Jr, R. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics,* 9(2), 508-523.

MacDonald, G., and M. S. Weisbach (2004). The Economics of Has-beens. *Journal of Political Economy,* 112(1), S289-S310.

Sampat, B. N. and A. A. Ziedonis (2004). Patent Citations and the Economic Value of Patents. In *Handbook of Quantitative Science and Technology Research*, 277-298. Springer.

Sarada, and Tocoian, O. (2015). Is It All About Who You Know? Prior Work Connections and Entrepreneurial Success. Duke University Working Paper.

Schumpeter, J. A. (1934). The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.*

Stokey, N. and R. Lucas Jr. (1989) *Recursive Methods in Economic Dynamics*, Harvard University Press.

Syverson, C (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2), 326–65.

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

Weinberg B. and D. Galenson (2005). Creative Careers: Lifecycles of Noble Laureates in Economics. National Bureau of Economic Research Working Paper 11799.

Weitzman, M. (1998). Recombinant Growth. *Quarterly Journal of Economics,* 113(2), 331-360.

Figures and Empirical Results

Table 1: Summary Statistics

Panel A: Descriptive Statistics

Variable	Observations	Mean	Standard Deviation
Unbalanced Firm Sample (Annual Firm Observations, 1992-2004)			
CEO age	7111	55.3	6.84
average manager age	7111	52.3	4.38
innovation quality	7111	15.9	10.9
superstar fraction	7111	9.91	10.7
tail innovation	7111	1.70	2.66
generality	6232	18.5	9.96
innovation quality (5 years)	4562	8.69	5.35
superstar fraction (best patent)	7111	33.0	20.0
tail innovation (99/50)	5803	3.41	5.42
originality	7091	27.1	8.49
log patents	7111	5.61	1.60
log employment	7111	3.71	1.51
log sales	7111	4.14	1.61
firm age	7111	35.1	16.3
profitability	7111	.009	1.28
indebtedness	7100	.776	.843
log physical capital	7091	7.96	1.68
R&D intensity	5922	.018	.087
CEO change probability (=1 if changed)	5387	.129	.335
CEO age change (unconditional)	5387	.004	4.49
CEO age change (conditional on CEO change)	693	-6.74	10.2
Change in innovation quality within firm over time	5387	-2.46	15.2
Change in superstar fraction within firm over time	5387	-1.17	13.6
Change in tail innovations within firm over time	5387	-.329	7.97
Change in generality within firm over time	4650	-2.31	14.7
Patent Sample (1992-2004)			
CEO age	316,516	55.7	6.77
inventor age	316,516	5.30	4.93
innovation quality	316,516	18.6	27.5
tail innovation (above 99)	316,516	18.7	136
tail innovation (above 90)	316,516	165	371
generality	263,461	21.9	24.0
log patents	316,516	5.63	1.59
log employment	316,516	3.73	1.49
log sales	316,516	4.15	1.60
inventor team size	316,516	2.59	1.74
maximum inventor patents	316,516	23.6	24.6

- Table 1 continued on next page -

Panel B: Correlation Matrix of Firm-Level Innovation Variables

	innovation quality	superstar fraction	tail innovation	generality
innovation quality	1.000			
superstar fraction	0.796	1.000		
tail innovation	0.608	0.591	1.000	
generality	0.435	0.140	-0.012	1.000

Panel C: Correlation Matrix of Patent-Level Innovation Variables

	innovation quality	tail innovation (above 99)	tail innovation (above 90)	generality
innovation quality	1.000			
tail innovation (above 99)	0.662	1.000		
tail innovation (above 90)	0.696	0.305	1.000	
generality	0.152	0.034	0.098	1.000

Notes: All statistics for the firm-level samples are weighted by the number of patents of the firm. Innovation quality is the average number of citations per patent (using the truncation correction weights devised by Hall, Jaffe, and Trajtenberg, 2001); superstar fraction is the fraction of patents accounted for by superstar researchers (those above the 95th percentile of the citation distribution); tail innovation is the fraction of patents of a firm above the 99th percentile of the citation distribution divided by all its patents; and the generality index measures the dispersion of citations received across two-digit IPC technology classes, whereas the originality index measures the dispersion of citations made by the patent to other patents. CEO age is the age of the CEO and average manager age is the average age of the top management, both from the Execucomp dataset. The unbalanced firm panel is a sample of firms from Compustat with at least one year of complete data between 1992 and 2004. Profitability is net income over sales. Indebtedness is total liabilities over sales. Physical capital is total net plant, property, and equipment. R&D intensity is R&D expenditures over sales, winsorized at the 99th percentile. See the text for the definition of other variables and further details.

Table 2: Baseline Cross-Sectional Regressions

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
CEO age	-0.168	-0.319	-0.078	-0.171
	(0.075)	(0.133)	(0.032)	(0.044)
firm age	-0.075	-0.102	-0.016	-0.017
	(0.025)	(0.036)	(0.007)	(0.017)
log employment	-1.208	-2.281	-0.386	-1.150
	(0.852)	(1.191)	(0.248)	(0.656)
log sales	1.477	2.071	0.254	1.292
	(0.812)	(1.107)	(0.229)	(0.574)
log patent	-0.367	0.002	0.099	-0.080
	(0.294)	(0.556)	(0.078)	(0.253)
N	7,111	7,111	7,111	6,232

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality. The key right-hand side variable is average CEO age (constant over time). Robust standard errors clustered at the firm level are in parentheses. A full set of four-digit SIC dummies, and year dummies (and thus no firm dummies) are included as controls. See text and notes to Table 1 for variable definitions.

Table 3: Cross-Sectional Regressions — Robustness

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
Panel A: Unweighted				
CEO age	-0.105	-0.199	-0.042	-0.041
	(0.056)	(0.073)	(0.024)	(0.047)
Panel B: Median Regression				
CEO age	-0.202	-0.467	-0.061	-0.071
	(0.085)	(0.078)	(0.004)	(0.010)
Panel C: No Covariates Except Time and SIC4 Fixed Effects				
CEO age	-0.166	-0.335	-0.085	-0.164
	(0.074)	(0.146)	(0.033)	(0.044)
Panel D: With SIC2 Dummies				
CEO age	-0.192	-0.371	-0.070	-0.010
	(0.044)	(0.084)	(0.021)	(0.076)
Panel E: With SIC3 Dummies				
CEO age	-0.188	-0.361	-0.077	-0.066
	(0.048)	(0.097)	(0.024)	(0.063)
Panel F: With Additional Controls				
CEO age	-0.167	-0.315	-0.077	-0.174
	(0.071)	(0.127)	(0.029)	(0.044)
Panel G: With Additional Controls Plus R&D Intensity				
CEO age	-0.173	-0.322	-0.080	-0.171
	(0.071)	(0.126)	(0.028)	(0.045)
R&D intensity	0.308	-2.422	0.981	1.137
	(2.183)	(2.353)	(0.776)	(1.707)
Panel H: Missing Observations as Zeroes				
CEO age	-0.115	-0.171	-0.034	-0.107
	(0.049)	(0.054)	(0.017)	(0.040)
Panel I: Controlling for Self-Citation Fraction				
CEO age	-0.116	-0.286	-0.062	-0.157
	(0.047)	(0.114)	(0.020)	(0.041)

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights unless stated otherwise. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality. The key right-hand side variable is average CEO age (constant over time). Each panel is for a different specification. Unless otherwise stated, all regressions control for firm age, log employment, log sales, log total patents, year dummies, and four-digit SIC dummies. Robust standard errors clustered at the firm level are in parentheses where applicable. Panel A reports the same regression in Table 2 without weights. Panel B runs median regressions following the same design as Table 2, while dropping weights and year and industry dummies. Panel C reports a regression without covariates except four-digit SIC dummies and year dummies. Panels D and E control for two- and three-digit SIC dummies respectively (instead of the four-digit SIC). Panel F adds to the specification of Table 2 profitability (profit over sales), indebtedness (debt over sales) and log physical capital. Panel G adds to the specification of Panel F R&D intensity (R&D expenditure over sales). Panel H replaces the missing values for dependent variables by zeroes, drops log patent count from the controls, and runs unweighted regressions in order to extend the sample to firms which do not apply for any patents in a given year. Panel I repeats the regression in Panel A adding self-citation fraction as a control. See text and notes to Table 1 for variable definitions.

Table 4: Cross-Sectional Regressions — Alternative Measures

	Innovation Quality (5 years)	Superstar Fraction (Best Patent)	Tail Innovation (90/0)	Originality
	Panel A: Weighted			
CEO age	-0.116	-0.596	-0.159	-0.278
	(0.047)	(0.337)	(0.064)	(0.103)
N	4,562	7,111	7,111	7,091
	Panel B: Unweighted			
CEO age	-0.056	-0.268	-0.139	-0.034
	(0.045)	(0.097)	(0.068)	(0.052)
N	4,562	7,111	7,111	7,091

	Tail Innovation (99/50)	Employment Growth	Sales Growth	R&D Intensity
	Panel C: Weighted			
CEO age	-0.104	-0.109	-0.144	-5.872
	(0.045)	(0.138)	(0.131)	(9.479)
N	5,803	5,387	5,387	5,922
	Panel D: Unweighted			
CEO age	-0.070	-0.165	-0.196	-46.148
	(0.041)	(0.081)	(0.090)	(86.190)
N	5,803	5,387	5,387	5,922

Notes: Firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables in Panels A and B are alternative measures of innovation quality (computed over the next five years), superstar fraction (with superstars defined according to the best patent), tail innovation (with share of the patents of the firm among all the patents above the 90th percentile of the citation distribution in the numerator), and the originality index. The dependent variables in Panels C and D are measures of tail innovation (with fraction of patents above the median in the denominator), employment growth and sales growth, and R&D intensity. The key right-hand side variable is average CEO age (constant over time). Robust standard errors clustered at the firm level are in parentheses. All regressions control for firm age, log employment, log sales, log total patents, year dummies and a full set of dummies for four-digit SIC industries. Regressions in Panel A and column 1 of Panel C are weighted by total patent count. Regressions in the last three columns of Panel C are weighted by firm employment. Regressions in Panels B and D are not weighted. See text and notes to Table 1 for variable definitions.

Table 5: Baseline Panel Regressions

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
Panel A: Time-varying CEO Age				
CEO age	-0.195	-0.259	-0.067	-0.070
	(0.045)	(0.069)	(0.019)	(0.034)
N	7,111	7,111	7,111	6,233
Panel B: CEO Age (Fixed Effects)				
CEO age	-0.188	-0.149	-0.048	0.036
	(0.044)	(0.051)	(0.012)	(0.029)
N	7,111	7,111	7,111	6,232
Panel C: CEO Age and Lagged CEO Age (Fixed Effects)				
CEO age	-0.131	-0.098	-0.035	0.031
	(0.041)	(0.039)	(0.013)	(0.026)
lagged CEO age	-0.123	-0.100	-0.029	0.020
	(0.051)	(0.049)	(0.016)	(0.035)
N	5,407	5,407	5,407	4,780
Panel D: CEO Age and Lagged Dependent Var (Fixed Effects)				
CEO age	-0.096	-0.075	-0.037	0.037
	(0.026)	(0.030)	(0.009)	(0.024)
lagged dependent variable	0.472	0.452	0.209	0.200
	(0.034)	(0.046)	(0.046)	(0.042)
N	5,985	5,985	5,985	5,207
Panel E: CEO Age and Lead CEO Age (Fixed Effects)				
CEO age	-0.113	-0.084	-0.028	0.042
	(0.042)	(0.048)	(0.011)	(0.029)
lead CEO age	-0.125	-0.109	-0.035	-0.007
	(0.049)	(0.044)	(0.014)	(0.028)
N	5,409	5,409	5,409	5,097

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality. Robust standard errors clustered at the firm level are in parentheses. All specifications control for log employment, log sales, log patents, year dummies and a full set of firm fixed effects (and thus firm age and the four-digit SIC dummies are no longer included). In Panel A, the key right-hand side variable is CEO age (in that year). Panel B is identical to Panel A except that it also includes a one year lag of CEO age as well as current CEO age. Panel C is identical to Panel A except that it also includes a one year lag of the dependent variable on the right-hand side. Panel D is identical to Panel A except that it also includes a one-year lead of the of CEO age in addition to current CEO age. See text and notes to Table 1 for variable definitions.

Table 6: Panel Regressions — Robustness

	Innovation Quality	Superstar Fraction	Tail Innovation	Innovation Quality	Superstar Fraction	Tail Innovation
			Panel A: Unweighted			
CEO age	-0.182	-0.135	-0.044	-0.175	-0.126	-0.037
	(0.050)	(0.041)	(0.023)	(0.109)	(0.044)	(0.036)
lead CEO age				-0.022	-0.009	0.010
				(0.126)	(0.058)	(0.043)
N	7,111	7,111	7,111	5,387	5,387	5,387
			Panel B: Missing Observations as Zeroes			
CEO age	-0.094	-0.083	-0.031	-0.092	-0.063	-0.029
	(0.035)	(0.031)	(0.013)	(0.063)	(0.036)	(0.020)
lead CEO age				-0.032	-0.034	-0.006
				(0.072)	(0.039)	(0.024)
N	11,525	11,525	11,525	10,009	10,009	10,009
			Panel C: High-Tech Subsample			
CEO age	-0.123	-0.083	-0.035	-0.048	-0.026	-0.003
	(0.054)	(0.082)	(0.015)	(0.053)	(0.084)	(0.017)
lead CEO age				-0.115	-0.096	-0.054
				(0.060)	(0.056)	(0.022)
N	2,089	2,089	2,089	1,698	1,698	1,698
			Panel D: Low-Tech Subsample			
CEO age	-0.223	-0.194	-0.057	-0.167	-0.141	-0.048
	(0.068)	(0.064)	(0.017)	(0.068)	(0.058)	(0.014)
lead CEO age				-0.110	-0.107	-0.020
				(0.069)	(0.057)	(0.018)
N	5,022	5,022	5,022	3,689	3,689	3,689
			Panel E: Controlling for Self-Citation Fraction			
CEO age	-0.185	-0.148	-0.048	-0.113	-0.088	-0.028
	(0.044)	(0.052)	(0.012)	(0.042)	(0.049)	(0.011)
lead CEO age				-0.119	-0.106	-0.033
				(0.048)	(0.044)	(0.014)
N	6,232	6,232	6,232	5,075	5,075	5,075

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, and tail innovation. Robust standard errors clustered at the firm level are in parentheses. All specifications control for log employment, log sales, log patents, year dummies and a full set of firm fixed effects (and thus firm age and the four-digit SIC dummies are no longer included). Panel A repeats the regressions in Table 5 Panels B and E without using weights. Panel B replaces the missing values for dependent variables by zeroes, drops log patent count from the controls, and runs unweighted regressions in order to extend the sample to firms which do not apply for any patents in a given year. Panels C and D are for the high-tech and low-tech subsamples. High-tech subsample includes all firms with a primary industry classification of SIC 35 (industrial and commercial machinery and equipment and computer equipment) and 36 (electronic and other electrical equipment and components), while the low-tech subsample includes the rest. Panel F repeats the same regressions adding self-citation fraction as a control. See text and notes to Table 1 for variable definitions.

Table 7: Structural Parameters

Parameter	*Description*	*Identification*
	External Calibration	
$x = 0.05$	Entry rate	Compustat sample
$\rho = 0.02$	Discount rate	Standard value
$\delta = 0.04$	Manager death rate	Compustat sample
$\alpha = 0.93$	Reduction rate of innovation size	Akcigit and Kerr (2015)
	Indirect Inference	
$\psi = 0.078$	Baseline radical innovation rate for high type	Estimate
$\Lambda\theta_H = 0.005$	High-type innovation parameter	Estimate
$\varphi = 0.040$	Transition rate from low type to high type	Estimate
$\xi = 0.037$	Incremental innovation rate	Estimate
$\eta = 0.433$	Initial innovation size	Estimate
$\zeta = 0.272$	Probability of high-type entrant	Estimate

Notes: Parameter choices and estimates. See Section 4.4 for details.

Table 8: Empirical and Model-Generated Moments

Target	*U.S. Data*	*Model*
Current manager age coefficient of Table 4 Panel C	-0.188	-0.189
Lead manager age coefficient of Table 4 Panel F	-0.125	-0.125
Current manager age coefficient of Table 4 Panel F	-0.113	-0.108
Annual growth rate	5.75%	5.95%
Within-firm coefficient of variation of radical innovations	1.99	2.23
Fraction of internal patents	21.5%	21.1%

Notes: Empirical and model-generated moments for the indirect inference procedure. See Section 4.4 for details.

Table 9: Patent-Level Panel Regressions

	Innovation Quality	Tail Innovation (Above 99)	Tail Innovation (Above 90)	Generality
Panel A: CEO Age, Unbalanced Firm Sample, 1992-2004				
CEO age	-0.119	-0.314	-1.239	0.028
	(0.038)	(0.132)	(0.413)	(0.025)
N	316,516	316,516	316,516	263,641
Panel B: Inventor Age, Unbalanced Firm Sample, 1992-2004				
inventor age	-0.234	-0.440	-2.883	-0.019
	(0.026)	(0.121)	(0.321)	(0.022)
N	316,516	316,516	316,516	263,641
Panel C: Inventor Age, Extended Sample, 1985-2004				
inventor age	-0.226	-0.377	-2.842	-0.017
	(0.022)	(0.075)	(0.293)	(0.017)
N	572,169	572,169	572,169	466,378
Panel D: Inventor Age, Extended Sample, 1985-2004				
inventor age	-0.201	-0.327	-2.359	-0.046
	(0.010)	(0.036)	(0.134)	(0.011)
N	1,855,887	1,855,887	1,855,887	1,550,825
Panel E: CEO Age and Inventor Age, Unbalanced Firm Sample, 1992-2004				
inventor age	-0.233	-0.438	-2.876	-0.019
	(0.026)	(0.121)	(0.321)	(0.022)
CEO age	-0.119	-0.317	-1.218	0.028
	(0.036)	(0.126)	(0.388)	(0.022)
N	316,516	316,516	316,516	263,641

Notes: Patent-level panel regressions with annual observations. The dependent variables are innovation quality at the patent level; a dummy for the patent being above the 99th percentile of the citation distribution; dummy for the patent being above the 90th percentile of the citation distribution; and generality index at the patent level. Robust standard errors clustered at the firm level are in parentheses. Panel A is for our unbalanced firm sample 1992-2004 and controls for log employment, log sales, log patents, a full set of firm fixed effects, and application year dummies, and the key right-and side variable is CEO age. Panel B is for our unbalanced firm sample 1992-2004 and controls for log employment, log sales, log patents, application year dummies, a full set of firm fixed effects, a full set of dummies for inventor team size, a full set of dummies for three-digit IPC technology class dummies, and a full set of dummies for the total number of patents of the inventor within the team with the highest number of patents, and the key right-and side variable is average inventor age. Panel C expands the sample of Panel B to 1985-2004 and also adds Compustat firms without CEO information into the sample. Panel D extends the sample of Panel C to include non-Compustat firms as well (hence excludes log sales and log employment, and still includes a full set of firm fixed effects). Panel E is for our unbalanced firm sample 1992-2004 and adds CEO age to the specification of Panel B. See text and notes to Table 1 for variable definitions.

Table 10: Inventor Age and CEO Age,
Unbalanced Firm Sample, 1992-2004

	Inventor age	Inventor age
	(1)	(2)
CEO age	0.014	0.013
	(0.006)	(0.006)
N	316,516	316,516

Notes: Patent-level panel regressions with annual observations for the unbalanced firm sample 1992-2002. The dependent variable is the average age of inventors. The first column controls for log employment, log sales, log patents, application year dummies, and a full set of firm fixed effects, and the second column adds to this a full set of team size dummies and a full set of dummies for three-digit IPC technology class dummies. See text and notes to Table 1 for variable definitions.

Table 11: Stock of Knowledge, Opportunity Cost, and Creative Innovations,
Unbalanced Firm Sample, 1992-2004

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
CEO age	-0.180	-0.216	-0.057	-0.044
	(0.027)	(0.027)	(0.008)	(0.016)
log sales	1.465	2.081	0.254	1.201
	(0.449)	(0.611)	(0.142)	(0.328)
log patent	-0.394	-0.072	0.082	-0.020
	(0.193)	(0.257)	(0.065)	(0.151)
CEO age $\times$ log patent	-0.005	-0.071	-0.016	-0.037
	(0.014)	(0.021)	(0.006)	(0.011)
CEO age $\times$ log sales	0.024	0.079	0.018	0.044
	(0.017)	(0.021)	(0.006)	(0.011)
N	7,111	7,111	7,111	6,232

Notes: Weighted firm-level panel regressions with annual observations for the unbalanced firm panel, 1992-2004, with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality. Robust standard errors clustered at the firm level are in parentheses. All regressions also include log employment, application year dummies and a full set of dummies for four-digit SIC industries. See text and notes to Table 1 for variable definitions.

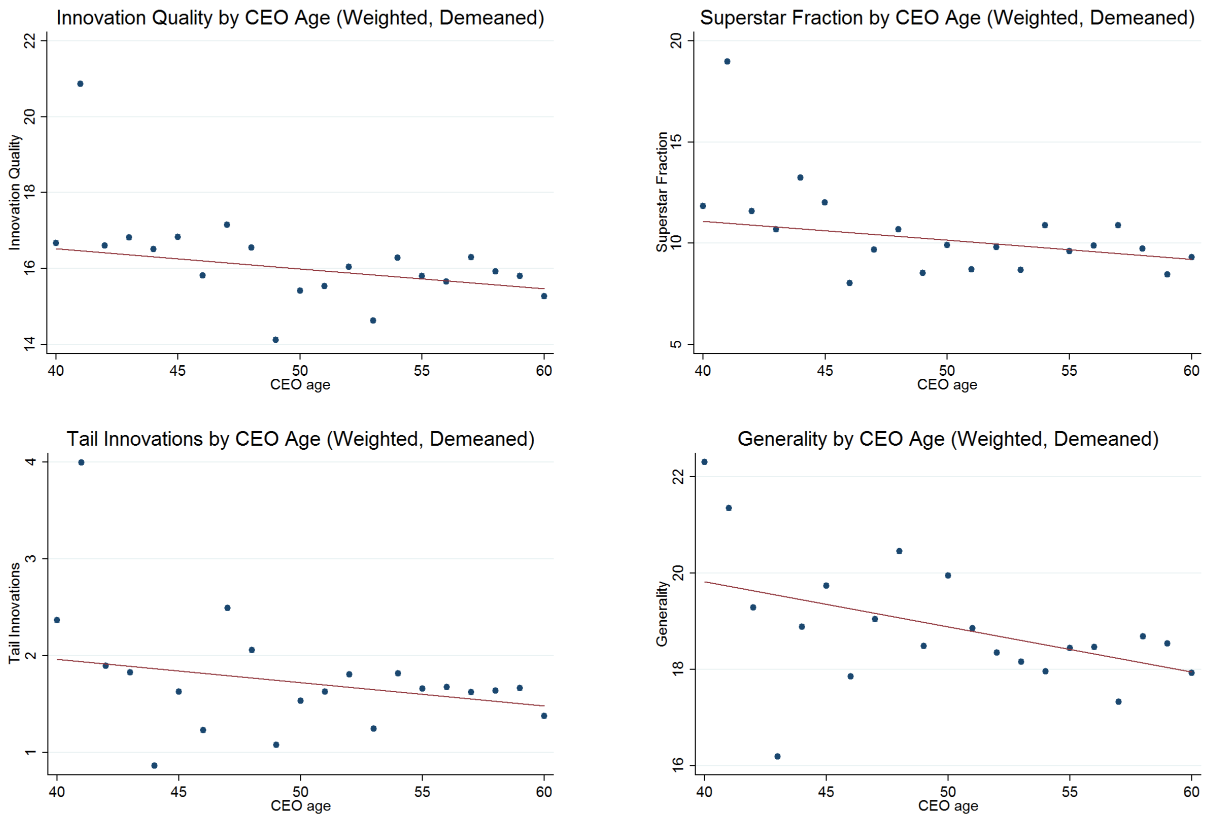


Figure 1: The averages for creative innovation variables (innovation quality, superstar fraction, tail innovations, and generality) by CEO age, and the associated fitted line. The averages are calculated after demeaning the creative innovation variables at the year × industry level. The label 40 stands for all ages less than or equal to 40, and 60 stands for all ages greater than or equal to 60. See text and notes to Table 1 for variable definitions.

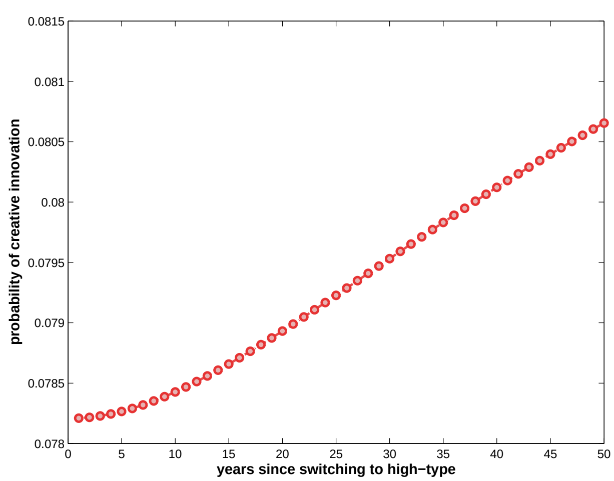


Figure 2: Evolution of creative innovations for high-type firms. The horizontal axis corresponds to the number of years since the firm has switched to high type (which is the year of entry for new high-type firms), and the vertical axis plots the average probability of generating a new radical innovation for all such firms.

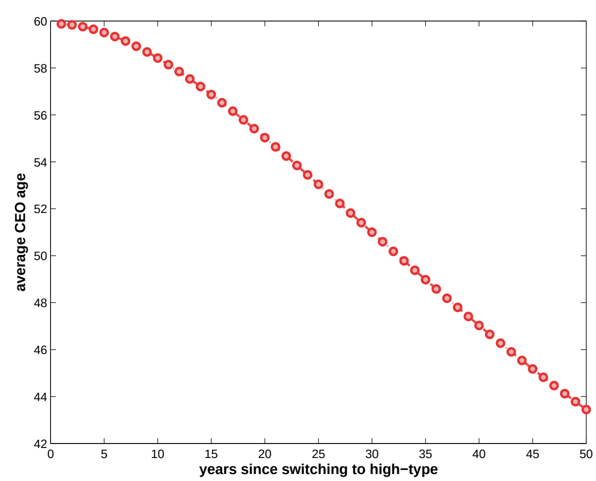


Figure 3: Evolution of CEO age for high-type firms. The horizontal axis corresponds to the number of years since the firm has switched to high type (which is the year of entry for new high-type firms), and the vertical axis plots the average age of the CEO for all such firms. Note that 26 has been added to the age of the CEO, since the youngest manager in the model has an age of zero, whereas this number is 26 in the data.

Appendix B: Online Materials for “Young, Restless and Creative: Openness to Disruption and Creative Innovations”

Daron Acemoglu Ufuk Akcigit Murat Alp Celik

November 1st, 2017

- Additional Graphs and Tables -

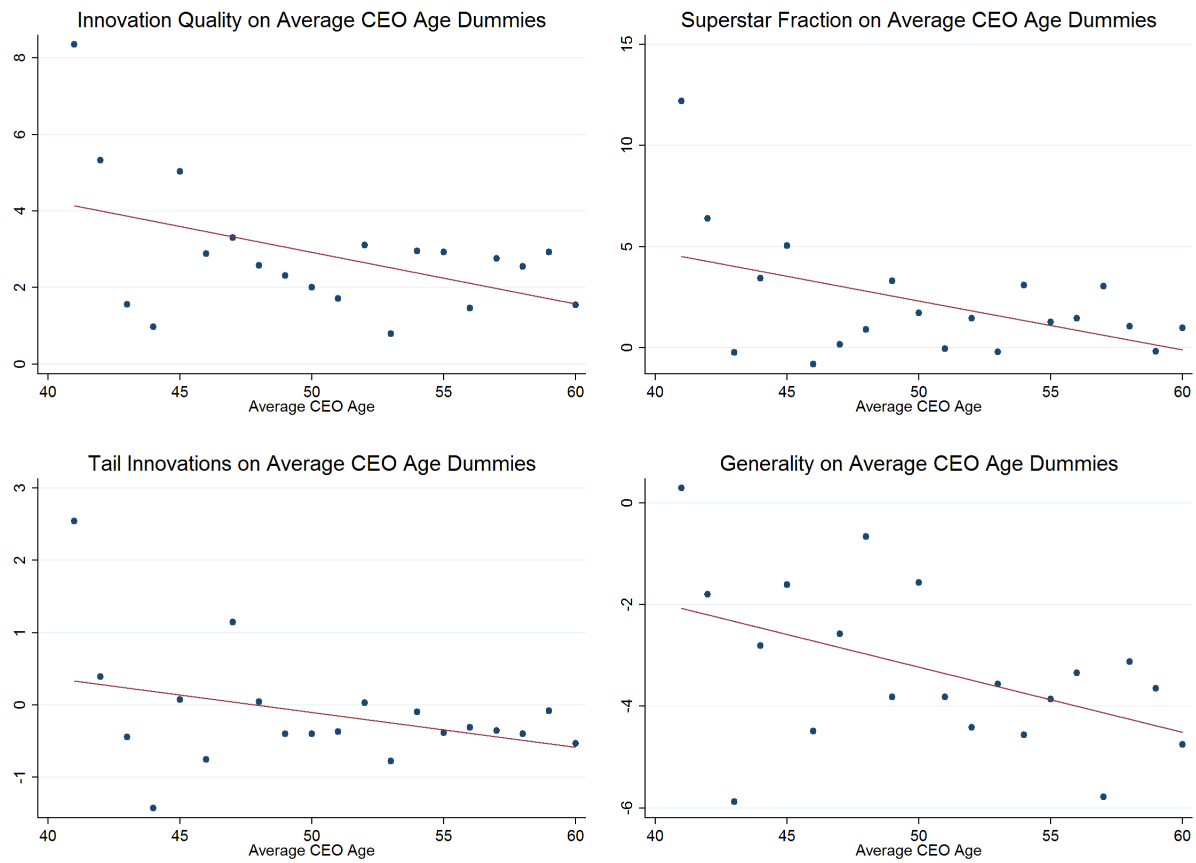


Figure B1: Estimated age dummies for the creative innovation variables (innovation quality, superstar fraction, tail innovations, and generality), and the associated fitted line. These figures repeat the regressions from Table 2, where CEO age is introduced as an array of discrete dummies instead of a linear regressor, and plot the estimated coefficients for the 21 age bins between ages of 40 and 60. The label 40 stands for all ages less than or equal to 40, and 60 stands for all ages greater than or equal to 60. See text and notes to Table 1 for variable definitions.

Table B1: Cross-Sectional Regressions — Further Robustness

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
Panel A: Balanced Sample				
CEO age	-0.278	-0.300	-0.117	-0.183
	(0.088)	(0.141)	(0.041)	(0.055)
N	279	279	279	279
Panel B: With Average Manager Age				
average manager age	-0.266	-0.536	-0.113	-0.208
	(0.118)	(0.185)	(0.050)	(0.080)
N	7,111	7,111	7,111	6,232
Panel C: High-Tech Subsample				
CEO age	-0.146	-0.273	-0.082	-0.180
	(0.096)	(0.157)	(0.036)	(0.050)
N	2,089	2,089	2,089	1,890
Panel D: Low-Tech Subsample				
CEO age	-0.232	-0.421	-0.062	-0.152
	(0.090)	(0.121)	(0.030)	(0.079)
N	5,022	5,022	5,022	4,342
Panel E: Non-Pharmaceuticals Subsample				
CEO age	-0.154	-0.306	-0.076	-0.174
	(0.073)	(0.132)	(0.032)	(0.045)
N	6,609	6,609	6,609	5,810

Notes: Weighted firm-level panel regressions (without fixed effects) with annual observations with number of patents (in that year) as weights unless stated otherwise. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality. The key right-hand side variable is average CEO age (constant over time). Each panel is for a different specification. Unless otherwise stated, all regressions control for firm age, log employment, log sales, log total patents, year dummies, and four-digit SIC dummies. Robust standard errors clustered at the firm level are in parentheses. Panel A displays a cross-sectional regression where all variables are the averages over the years 1995-2000 for a balanced sample of 279 firms. Panel B uses average manager age instead of CEO age. Panels C and D are for the high-tech and low-tech subsamples. High-tech subsample includes all firms with a primary industry classification of SIC 35 (industrial and commercial machinery and equipment and computer equipment) and 36 (electronic and other electrical equipment and components), while the low-tech subsample includes the rest. Panel E repeats the regression on Table 2 while dropping the pharmaceutical sector from the sample (SIC 283). See text and notes to Table 1 for variable definitions.

Table B2: Panel Regressions (with Fixed Effects) — Further Robustness

	Innovation Quality	Superstar Fraction	Tail Innovation	Innovation Quality	Superstar Fraction	Tail Innovation
Panel A: No Covariates Except Time and Firm Fixed Effects						
CEO age	-0.243	-0.192	-0.059	-0.160	-0.123	-0.037
	(0.059)	(0.063)	(0.014)	(0.051)	(0.056)	(0.010)
lead CEO age				-0.160	-0.135	-0.041
				(0.055)	(0.046)	(0.016)
N	7,111	7,111	7,111	5,387	5,387	5,387
Panel B: With Additional Controls						
CEO age	-0.187	-0.149	-0.049	-0.116	-0.089	-0.029
	(0.045)	(0.051)	(0.012)	(0.044)	(0.049)	(0.011)
lead CEO age				-0.124	-0.111	-0.035
				(0.048)	(0.044)	(0.014)
N	7,080	7,080	7,080	5,372	5,372	5,372
Panel C: With Additional Controls Plus R&D Intensity						
CEO age	-0.185	-0.147	-0.048	-0.115	-0.086	-0.029
	(0.046)	(0.052)	(0.012)	(0.044)	(0.050)	(0.011)
lead CEO age				-0.124	-0.113	-0.035
				(0.049)	(0.044)	(0.014)
R&D intensity	0.145	-1.911	1.062	1.614	-2.258	1.446
	(2.839)	(2.243)	(0.932)	(3.277)	(2.951)	(1.054)
N	5,907	5,907	5,907	4,699	4,699	4,699
Panel D: Non-Pharmaceuticals Subsample						
CEO age	-0.188	-0.152	-0.049	-0.107	-0.085	-0.028
	(0.046)	(0.053)	(0.013)	(0.044)	(0.051)	(0.012)
lead CEO age				-0.143	-0.124	-0.038
				(0.051)	(0.046)	(0.015)
N	6,609	6,609	6,609	4,986	4,986	4,986
Panel E: Median Regression						
CEO age	-0.145	-0.095	-0.021	-0.128	-0.061	-0.018
	(0.028)	(0.031)	(0.007)	(0.014)	(0.054)	(0.007)
lead CEO age				-0.049	-0.066	-0.008
				(0.013)	(0.051)	(0.007)
N	7,111	7,111	7,111	5,387	5,387	5,387

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, and tail innovation. Robust standard errors clustered at the firm level are in parentheses. All specifications control for log employment, log sales, log patents, year dummies and a full set of firm fixed effects unless mentioned otherwise. Panel A repeats the regressions in Table 5 Panels B and E where all controls except year and firm fixed effects are dropped. Panel B repeats the same regressions while introducing the additional controls profitability, indebtedness and log physical capital. Panel C repeats the same regression as Panel B with the addition of R&D intensity as a control. Panel D repeats the same regressions while dropping the pharmaceuticals sector from the sample (SIC 283). Finally, Panel E reports the results of a median regression, where we first demean all observations to remove firm fixed effects. See text and notes to Table 1 for variable definitions.

Table B3: Panel Regressions (with Fixed Effects) — Alternative Measures

	Innovation Quality (5 years)	Superstar Fraction (Best Patent)	Tail Innovation (90/0)	Originality
	Panel A: Weighted			
CEO age	-0.066	-0.109	-0.211	-0.032
	(0.039)	(0.063)	(0.052)	(0.027)
N	4,562	7,111	7,111	7,091
	Panel B: Unweighted			
CEO age	-0.077	-0.088	-0.239	-0.038
	(0.056)	(0.053)	(0.057)	(0.041)
N	4,562	7,111	7,111	7,091

	Tail Innovation (99/50)	Employment Growth	Sales Growth	R&D Intensity
	Panel C: Weighted			
CEO age	-0.076	0.135	0.016	0.857
	(0.023)	(0.198)	(0.126)	(3.223)
N	5,803	5,387	5,387	5,922
	Panel D: Unweighted			
CEO age	-0.046	-0.323	0.045	154.371
	(0.036)	(0.296)	(0.210)	(120.215)
N	5,803	5,387	5,387	5,922

Notes: Firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables in Panels A and B are alternative measures of innovation quality (computed over the next five years), superstar fraction (with superstars defined according to the best patent), tail innovation (with share of the patents of the firm among all the patents above the 90th percentile of the citation distribution in the numerator), and the originality index. The dependent variables in Panels C and D are measures of tail innovation (with fraction of patents above the median in the denominator), employment growth and sales growth, and R&D intensity. The key right-hand side variable is CEO age. Robust standard errors clustered at the firm level are in parentheses. All regressions control for log employment, log sales, log total patents, year and firm dummies. Regressions in Panel A and column 1 of Panel C are weighted by total patent count. Regressions in columns 2-4 of Panel C are weighted by firm employment. Regressions in Panels B and D are not weighted. See text and notes to Table 1 for variable definitions.

Table B4:Cross-Sectional and Panel Regressions Excluding Self-Citations

	Innovation Quality (Const. age)	Tail Innovation (Const. age)	Innovation Quality (Firm F.E.)	Tail Innovation (Firm F.E.)	Innovation Quality (Firm F.E.)	Tail Innovation (Firm F.E.)
CEO age	0.084	0.014	-0.154	-0.027	-0.095	-0.015
	(0.062)	(0.015)	(0.040)	(0.008)	(0.035)	(0.008)
lead CEO age					-0.101	-0.018
					(0.043)	(0.010)
N	7,111	7,111	7,111	7,111	5,387	5,387

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are innovation quality and tail innovation which are calculated after excluding all self-citations (citations made by patents that belong to the same firm). Robust standard errors clustered at the firm level are in parentheses. The first two columns control for firm age, log employment, log sales, log patents, year dummies, and a full set of four-digit SIC industry dummies. Columns 3-6 control for log employment, log sales, log patents, year dummies and a full set of firm fixed effects (and thus firm age and the four-digit SIC dummies are no longer included). The key right-hand side variable is average CEO age (constant over time) for the first two columns, CEO age for columns 3 and 4, and current and lead CEO age for columns 5 and 6. See text and notes to Table 1 for variable definitions.

Table B5: Cross-Sectional and Panel Regressions Separating Citations by Source

	All Citations	Examiner Citations Excluded	Self and Examiner Citations Excluded
	Panel A: With Average CEO Age		
CEO age	-0.058	-0.066	-0.058
	(0.052)	(0.047)	(0.046)
N	1,610	1,610	1,610
	Panel B: With Firm Fixed Effects		
CEO age	-0.052	-0.039	-0.040
	(0.061)	(0.057)	(0.056)
N	1,610	1,610	1,610

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are citations received within five years with different exclusion restrictions. The first column includes all citations. The second column excludes citations that applicants make to their own patents. The third column excludes citations added by patent examiners. The fourth column excludes both citations added by patent examiners and self-citations. The key right-hand side variable is average CEO age (constant over time) in Panel A, and CEO age (time-varying) in Panel B. Robust standard errors clustered at the firm level are in parentheses. Panel A controls include firm age, log sales, log employment, log total patents, year and four-digit SIC dummies. Panel B controls include log sales, log employment, log total patents, year dummies, and a full set of firm fixed effects.

Table B6: Continuing Inventors vs New Hires — Innovation Quality

	Innovation Quality (All)	Innovation Quality (New Inventors)	Innovation Quality (Continuing Inventors)
CEO age	-0.188	-0.205	-0.159
	(0.044)	(0.045)	(0.042)
N	7,111	5,769	5,541
mean of dep. var.	15.9	16.8	17.3

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables is the innovation quality measure calculated in three different ways. The first column uses the information from all the patents of a firm. The second column uses the information only from patents created by new inventors, where a new inventor is defined as an inventor who has never worked for the particular firm before. The third column uses the information only from patents created by continuing inventors. When a patent is created by a mix of new and continuing inventors, it is weighted according to the fraction of new vs. continuing inventors, defined as the inverse of a new inventor. The key right-hand side variable is CEO age. Robust standard errors clustered at the firm level are in parentheses. All specifications control for log employment, log sales, log patents, year dummies and a full set of firm fixed effects (and thus firm age and the four-digit SIC dummies are no longer included). See text and notes to Table 1 for variable definitions.

Table B7: Continuing Inventors vs New Hires — Tail Innovations

	Tail Innovations (All)	Tail Innovations (New Inventors)	Tail Innovations (Continuing Inventors)
CEO age	-0.048	-0.051	-0.034
	(0.012)	(0.020)	(0.023)
N	7,111	5,769	5,541
mean of dep. var	1.70	1.89	2.05

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables is the tail innovation measure calculated in three different ways. The first column uses the information from all the patents of a firm. The second column uses the information only from patents created by new inventors, where a new inventor is defined as an inventor who has never worked for the particular firm before. The third column uses the information only from patents created by continuing inventors. When a patent is created by a mix of new and continuing inventors, it is weighted according to the fraction of new vs. continuing inventors, defined as the inverse of a new inventor. The key right-hand side variable is CEO age. Robust standard errors clustered at the firm level are in parentheses. All specifications control for log employment, log sales, log patents, year dummies and a full set of firm fixed effects (and thus firm age and the four-digit SIC dummies are no longer included). See text and notes to Table 1 for variable definitions.