American Economic Journal: Macroeconomics 2022, 14(3): 199–249
https://doi.org/10.1257/mac.20170410

Radical and Incremental Innovation: The Roles of Firms, Managers, and Innovators†

By Daron Acemoglu, Ufuk Akcigit, and Murat Alp Celik*

We investigate the determinants of radical (“creative”) innovations that break new ground in knowledge creation. We develop a model focusing on the choice between incremental and radical innovation and on how managers of different ages and human capital are sorted across firms. Firm- and patent-level evidence reveals that firms that are more “open to disruption” are significantly more likely to engage in radical innovation and hire younger managers and inventors with a comparative advantage in radical innovation. However, once the effect of the sorting is factored in, the (causal) impact of manager age on creative innovations, though positive, is small. (*JEL* D22, L26, M10, M14, O31, O34)

Radical (creative) innovations play an important role in economic growth not just because of their direct contributions to productivity but also because further innovations can build on them. 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 217,001 patents between 1976 and 2001, the median number of citations within the next 5 years was 3 (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

* Acemoglu: Massachusetts Institute of Technology (email: daron@mit.edu); Akcigit: University of Chicago (email: uakcigit@uchicago.edu); Celik: University of Toronto (email: murat.celik@utoronto.ca). Richard Rogerson was coeditor for this article. 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 University, MIT, University of Pennsylvania, AEA Meetings, SED 2014, 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 and NSF 1654719), 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 (Grant 0735956) is gratefully acknowledged.

† Go to https://doi.org/10.1257/mac.20170410 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] See, among others, Trajtenberg (1990); Harhoff et al. (1999); and Sampat and Ziedonis (2005) on the relationship between citations and patent quality.

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 5 years. Similarly, in computer and communications, the median number of citations within the next 5 years in this period is 6, 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 5 years.

Where do radical innovations come from? In this paper, we investigate one aspect of this question, focusing on the roles of firms, managers, and inventors, and the contribution of younger managers and inventors and their sorting to firms that have a comparative advantage in radical innovations.

We first provide a simple model of the interplay between firms that are heterogeneous in terms of their ability to undertake radical innovation and managers. In our model, all firms can engage in *incremental innovation* by building on their existing leading-edge products. In addition, high-type firms can also attempt a *radical innovation*, which involves combining diverse ideas to generate a technological improvement in a new area. We interpret high-type firms as those with a "corporate culture" that is open to radical ideas and disruption, though there may be other aspects that make some firms more successful in radical innovations.[2] We also 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 2018 or Abrams, Akcigit, and Grennan 2018), 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 radical innovations because high-type firms tend to hire young managers, and thus such cross-sectional relationships also capture the "sorting" channel. Indeed, in our model, young managers sort to firms that are both high type and willing to undertake radical innovations.[3] These forces can be seen from the longitudinal predictions of the model: firms that hire younger managers should subsequently have more radical 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

[2] A natural interpretation (which we favor) is to identify these high-type firms with Joseph Schumpeter's (1934) "creative agents" (firms, managers, entrepreneurs, and inventors) that are open to disruptions and have the "mental freedom" to deviate from existing technologies, practices, and rules of organizations and societies in order to engage in "disruptive innovations" (pp. 86–94). Nevertheless, there are other aspects that make some firms have a comparative advantage in radical innovations, and our model does not take a position on the exact source of this difference.

[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 only 41 at the time, and 43 at Sun Microsystem (compared to an average age of 51 among Compustat companies).

because high-type firms do not immediately hire a young manager and switch to radical innovations, the increase in radical innovations typically precedes 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 that build on their existing product lines, and those with many patents will tilt things in favor of radical innovations (because of diminishing returns).

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 high-type firms (or those with greater openness to disruption), engage in more radical and creative innovations. Our empirical work uses several different measures of radical 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 ninety-ninth 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.

Our empirical results provide nuanced support for the role of firms, managers, and innovators in radical innovations. On the one hand, we establish a robust cross-sectional correlation between CEO age and all of our measures of firm-level radical 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.

On the other hand, our results suggest that much of this relationship is due to the greater innovativeness of high-type firms (for example, because of their greater openness to destruction and new ideas), while younger managers have a positive but much smaller effect. First, consistent with our theory, we find that firms switch to radical innovation even before they hire a younger manager. Second, when we use the structure of our model in conjunction with the reduced-form patterns in the data to estimate the causal effect of young managers on radical innovation, our estimates are typically small—younger managers accounting for about 1–7 percent of the total amount of radical innovation in the economy.

We further shed light on the role of innovators in the innovation process using the patent-level variation. Our estimates here indicate that younger CEOs tend to work with younger inventors and that younger inventors are significantly more creative and likely to generate radical innovations.

Finally, we investigate our model's prediction that firms with greater sales and with fewer patents should be less willing to engage in radical 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 2018; Hurst and Pugsley 2011; Syverson 2011; Kogan et al. 2017; Acemoglu et al. 2018) 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 (2017); and Fogli and Veldkamp (2021) 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 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 (2000, 2001); Weinberg and Galenson (2019); Jones and Weinberg

[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).

(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 (2015), 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, Pérez-González, and Wolfenzon (2020) 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).[7]

The rest of the paper is organized as follows. Section I presents our motivating model. Section II describes our data. Section III presents our main empirical results. Section IV concludes, while Appendix A contains the proofs from Section I and some supplementary materials. Online Appendix B, which is not for publication, presents additional figures, empirical results, and discusses the possible microfoundations of the critical assumptions we make in the theoretical model.

I. 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. Further discussion of assumptions and microfoundations of the model are provided in online Appendix B1.

A. *Production*

We consider a continuous-time economy in which discounted preferences are defined over a unique final good as $\int_0^\infty e^{-\rho t}((C(t)^{1-\nu}-1)/(1-\nu))\,dt$,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 labor and a

[6] See also Sarada and Tocoian (2019), who investigate the impact of the age of the founders of a company on subsequent performance using Brazilian data; Azoulay, Zivin, and Manso (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 coauthor 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).

[7] See also Bandiera et al. (2020), who use CEO diary data and machine learning techniques to differentiate between "leader" CEOs (i.e., those primarily involved in communication and coordination activities) and "manager" CEOs (i.e., those primarily involved in production-related activities). They show that firms with leader CEOs are larger and more productive on average.

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 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

$$p_j = L^{\beta} q_j^{\beta} k_j^{-\beta}, \quad \forall j \in \mathcal{C}. \tag{2}$$

The profit-maximization problem of the firm with the leading-edge technology for intermediate j can then be written as

$$\Pi(q_j) = \max_{k_j \geq 0} \{p_j k_j - \gamma k_j\} \; \forall j \in \mathcal{C} \quad \text{subject to } (2).$$

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

$$k_j = \left[(1-\beta)/\gamma\right]^{\frac{1}{\beta}} L q_j. \tag{3}$$

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

$$\bar{q}_t \equiv \int_{\mathcal{C}} q_{jt} dj. \tag{4}$$

B. *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 $1 - e^{-\delta a}$.[8]

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

$$\bar{q}_b \equiv \int_C q_{jb}\, dj.$$

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).

C. *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 that have a comparative advantage in radical innovations, for example, because they have corporate cultures that are open to disruption. 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 $\varphi \in (0,1)$.[9]

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.

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

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

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 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 nth 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

$$\eta_n(q_j, \bar{q}_t) = \left[\kappa \bar{q}_t + (1 - \kappa) q_j\right] \eta \alpha^n, \tag{5}$$

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.[10] 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

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

[10] 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).[11] 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

$$\Lambda\theta\bar{q}^a, \tag{6}$$

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, (6) 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.[12]

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

[11] 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.

[12] 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.

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.

D. *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. 2018). More formally, let us define $W_s(\vec{q}_f, \vec{n}_f)$ (for $s \in \{H, L\}$) as the value of a firm with a vector of products $\vec{q}_f = \{q_{f,j_1}, q_{f,j_2}, \ldots, q_{f,j_{m_f}}\}$, with associated number of incremental innovations $\vec{n}_f = \{n_{f,j_1}, n_{f,j_2}, \ldots, n_{f,j_{m_f}}\}$.[13] Thanks to this independence property, as we show in Appendix A, firm values satisfy

$$W_s(\vec{q}_f, \vec{n}_f) = \sum_{m=1}^{m_f} V_s(q_j, n), \tag{7}$$

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

$$\begin{aligned} rV_L(q_j,n) - \dot{V}_L(q_j,n) = {} & \max_a\{\pi q_j + \bar{q}_t f(a) - w_{a,t}\} \\ & + \xi\left[V_L(q_j + \eta_{n+1}, n+1) - V_L(q_j,n)\right] \\ & - \tau V_L(q_j,n) + \varphi\left[V_H(q_j,n) - V_L(q_j,n)\right], \end{aligned} \tag{8}$$

and

$$\begin{aligned} & rV_H(q_j,n) - \dot{V}_H(q_j,n) \\ & \quad = \max\Big\{\pi q_j + \max_a\Big\{\bar{q}_t f(a) - w_{a,t} \\ & \qquad\qquad + \xi\left[V_H(q_j + \eta_{n+1}, n+1) - V_H(q_j,n)\right]\Big\}; \\ & \qquad\qquad \pi q_j + \max_a\left\{\bar{q}_t f(a) + \Lambda\theta_H \bar{q}^a EV_H(\bar{q}_t) - w_{a,t}\right\}\Big\} \\ & \qquad - \tau V_H(q_j,n) + \psi\Lambda\theta_H EV_H(\bar{q}_t). \end{aligned} \tag{9}$$

Here r is the equilibrium interest rate, τ is the rate of creative destruction in the economy, and $EV_H(\bar{q}_t)$ 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 (8), a low-type firm's value from

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

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 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 (8) 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 (9) is similar except that it involves an additional choice between incremental innovation and radical innovation, as represented by the inner maximization.

E. *Stationary Equilibrium with* $\kappa = 1$

We now characterize the stationary equilibrium of this economy. We start with the case where $\kappa = 1$ in equation (5)—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

$$w_{a,t} = \bar{q}_t f(a). \tag{10}$$

We next turn to the value of a product line operated by a high-type firm, (9). Because of the comparative advantage of young managers for radical innovation in (6), 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 (9) implies that firms engaged in radical innovation must be indifferent between hiring any manager younger than a^*, and thus,

$$\bar{q}_t f(a^*) + \Lambda \theta_H \bar{q}^{a^*} E V_H(\bar{q}_t) - w_{a^*,t} = \bar{q}_t f(a) + \Lambda \theta_H \bar{q}^{a} E V_H(\bar{q}_t) - w_{a,t}$$

for all

$$a < a^*.$$

The boundary condition, (10), implies that the oldest manager hired for radical innovation earns $w_{a^*,t} = \bar{q}_t f(a^*)$. Hence,

$$w_{a,t} = \begin{cases} \bar{q}_t f(a) & \text{for} \quad a \geq a^* \\ \bar{q}_t f(a) + \Lambda \theta_H [\bar{q}^a - \bar{q}^{a^*}] E V_H(\bar{q}_t) & \text{for} \quad a < a^* \end{cases}. \tag{11}$$

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.[14]

The next proposition provides the characterization of the stationary equilibrium. It is important to note that 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.

PROPOSITION 1: *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:*

(*i*) *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 $b < b_t^*$);*

(*ii*) *they pursue radical innovations, $\lambda_n = 1$, on product lines with $n > n^*$; and*

(*iii*) *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 $\lambda_{n^*} \in [0,1]$.

[14]The evidence in Galenson and Weinberg (2000, 2001); Weinberg and Galenson (2019); and Jones and Weinberg (2011) is consistent with the possibility that either younger or older creative workers might be more productive.

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 radical innovations. We will first investigate the cross-sectional relationship between manager (CEO) age and radical innovations. In these cross-sectional regressions, manager age is taken to be a proxy for high-type firms (for example, those with a corporate culture that is more open to disruption). Therefore, from Proposition 1, we expect a negative cross-sectional relationship between manager age and radical innovations. As just stressed, this cross-sectional relationship does not correspond to the "causal effect" of manager age on creativity 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 high-type firms.

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

$$\psi \Lambda \theta_H + \frac{1}{F(a^*)} \int_0^{a^*} \Lambda \theta_H \bar{q}^a \, dF(a) = \psi \Lambda \theta_H + \frac{\Lambda \theta_H \delta}{g + \delta} \frac{\left[1 - e^{-(g+\delta)a^*}\right]}{\left[1 - e^{-\delta a^*}\right]}. \tag{12}$$

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.[15] 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 (12) times λ_{n^*}. In contrast, if the product

[15] 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.

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 (12).

This discussion clarifies that when we focus on the relationship between within-firm changes in manager age and radical 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 radical innovations. Second, firms that switch from an older to a younger manager should, on average, experience an increase in radical 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 radical 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 IIIC 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.

F. *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.[16]

II. 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.

A. *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 3,210,361 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.[17]

Compustat North American Fundamentals.—We draw our main sample from the Compustat (Standard and Poor's 2012) for publicly traded firms in North America. This dataset contains balance sheets reported by the companies annually between 1950 and 2012. 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 comparative advantage for radical innovations or openness to disruption at the firm level.[18]

The Careers and Coauthorship Networks of US Patent Inventors.—Extensive information on the inventors of patents granted in the United States between years 1975 and 2008 is obtained from Lai et al.'s (2014) 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.

[16] 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.

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

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

B. *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 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 percent 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 pth 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 ninety-ninth 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.

C. *Sample and Descriptive Statistics*

Our baseline analysis focuses on an unbalanced *firm sample* comprised of 7,170 observations from 1,259 firms between 1992 and 2004. Our data on CEO age do not

Table 1—Summary Statistics

Variable	Observations	Mean	Standard deviation
Panel A. Descriptive statistics			
Unbalanced firm sample (annual firm observations, 1992–2004)			
CEO age	7,170	55.3	6.85
Average manager age	7,170	52.2	4.38
Innovation quality	7,170	15.9	10.9
Superstar fraction	7,170	9.94	10.8
Tail innovation	7,170	1.71	2.67
Generality	6,286	18.5	9.95
Innovation quality (5 years)	4,606	8.70	5.36
Superstar fraction (best patent)	7,170	33.0	20.0
Tail innovation (99/50)	5,852	3.42	5.41
Originality	7,150	27.1	8.49
log patents	7,170	5.61	1.60
log employment	7,170	3.71	1.51
log sales	7,170	4.14	1.61
Firm age	7,170	35.0	16.3
Profitability	7,170	0.009	1.28
Indebtedness	7,159	0.775	0.843
log physical capital	7,150	7.95	1.68
R&D intensity	5,966	0.018	0.087
CEO change probability (=1 if changed)	5,472	0.128	0.334
CEO age change (unconditional)	5,472	0.011	4.47
CEO age change (conditional on CEO change)	701	−6.72	10.2
Change in innovation quality within firm over time	5,472	−2.48	15.2
Change in superstar fraction within firm over time	5,472	−1.19	13.6
Change in tail innovations within firm over time	5,472	−0.328	7.91
Change in generality within firm over time	4,731	−2.30	14.8
Patent Sample (1992–2004)			
CEO age	318,007	55.7	6.77
Inventor age	318,007	5.29	4.94
Innovation quality	318,007	18.6	27.6
Tail innovation (above 99)	318,007	18.8	136
Tail innovation (above 90)	318,007	165	371
Generality	264,972	21.9	24.0
log patents	318,007	5.63	1.59
log employment	318,007	3.73	1.49
log sales	318,007	4.15	1.60
Inventor team size	318,007	2.59	1.74
Maximum inventor patents	318,007	23.6	24.6

(*continued*)

extend before 1992, and we cannot go further than 2004 since our patent citation data end in 2006 and we need at least two postgrant years for citation analysis. We also study a larger *patent sample* of 318,007 patents (from 1,195 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, and there are also substantial variations (standard deviations are 6.8 in both samples). 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.

Table 1—Summary Statistics (*continued*)

	Innovation quality	Superstar fraction	Tail innovation	Generality
Panel B. Correlation matrix of firm-level innovation variables				
Innovation quality	1.000			
Superstar fraction	0.796	1.000		
Tail innovation	0.609	0.593	1.000	
Generality	0.435	0.141	−0.010	1.000

	Innovation quality	Tail innovation (above 99)	Tail innovation (above 90)	Generality
Panel C. Correlation matrix of patent-level innovation variables				
Innovation quality	1.000			
Tail innovation (above 99)	0.663	1.000		
Tail innovation (above 90)	0.696	0.306	1.000	
Generality	0.152	0.035	0.099	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 ninety-fifth percentile of the citation distribution); tail innovation is the fraction of patents of a firm above the ninety-ninth 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 ninety-ninth percentile. See the text for the definition of other variables and further details.

III. Empirical Results

In our theory, manager age is partly an indicator of a corporate culture that is open to disruption (because high-type firms have a comparative advantage in radical innovation and so tend to select younger managers) but also has a causal effect on radical 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 radical innovations and manager age.[19] We then turn to a more direct investigation of the effect of manager age on radical innovations, focusing on "within-firm" variation. Finally, exploiting the timing of the increase in radical innovations following a change in manager age, we provide estimates of the structural parameters of the model.

A. *Cross-Sectional Results*

Our cross-sectional results are presented in Tables 2 and 3, online Appendix Table B6, and in Figure 1. Our estimating equation is

$$y_f = \alpha m_f + \mathbf{X}_f'\beta + \delta_{i(f)} + \nu_t + \varepsilon_f, \tag{13}$$

[19] 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).

Table 2—Baseline Cross-Sectional Regressions

	Innovation quality	Superstar fraction	Tail innovation	Generality
CEO age	−0.171	−0.320	−0.079	−0.172
	(0.075)	(0.132)	(0.031)	(0.044)
Firm age	−0.075	−0.102	−0.015	−0.017
	(0.025)	(0.036)	(0.007)	(0.017)
log employment	−1.172	−2.246	−0.385	−1.147
	(0.852)	(1.189)	(0.248)	(0.654)
log sales	1.453	2.044	0.252	1.281
	(0.813)	(1.105)	(0.229)	(0.573)
log patent	−0.364	0.008	0.102	−0.067
	(0.294)	(0.554)	(0.078)	(0.253)
Observations	7,170	7,170	7,170	6,286

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.

where y_f is one of our measures of radical 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 comparative advantage in radical innovation or 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. 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,[20] and ν_t denotes a full set of 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.[21] We turn to specifications with fixed effects in Table 4.

Unless otherwise indicated, all of our regressions have one observation per firm $\times$ 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 radical 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 radical 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, -0.171 (standard error $=$ 0.075), is statistically significant at less than 5 percent and indicates that companies with a younger CEO have greater innovation quality. We interpret this pattern as evidence that companies that

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

[21] Panel A of Table 4 shows very similar, but more precisely estimated, results when the CEO age variable is time-varying (without fixed effects).

Table 3—Cross-Sectional Regressions—Robustness

	Innovation quality	Superstar fraction	Tail innovation	Generality
Panel A. Unweighted				
CEO age	−0.108	−0.194	−0.042	−0.045
	(0.056)	(0.073)	(0.024)	(0.047)
Panel B. Median regression				
CEO age	−0.200	−0.482	−0.063	0.063
	(0.080)	(0.070)	(0.005)	(0.007)
Panel C. No covariates except time and SIC4 fixed effects				
CEO age	−0.168	−0.336	−0.086	−0.162
	(0.073)	(0.146)	(0.033)	(0.045)
Panel D. With SIC2 dummies				
CEO age	−0.194	−0.374	−0.071	0.009
	(0.044)	(0.084)	(0.021)	(0.076)
Panel E. With SIC3 dummies				
CEO age	−0.191	−0.364	−0.077	−0.068
	(0.048)	(0.096)	(0.024)	(0.063)
Panel F. With additional controls				
CEO age	−0.170	−0.317	−0.077	−0.175
	(0.070)	(0.126)	(0.028)	(0.043)
Panel G. With additional controls plus R&D intensity				
CEO age	−0.175	−0.324	−0.081	−0.171
	(0.071)	(0.126)	(0.028)	(0.045)
R&D intensity	0.341	−2.351	1.001	1.118
	(2.177)	(2.334)	(0.769)	(1.709)
Panel H. Missing observations as zeros				
CEO age	−0.115	−0.168	−0.034	−0.104
	(0.049)	(0.054)	(0.017)	(0.040)
Panel I. Controlling for self-citation fraction				
CEO age	−0.118	−0.287	−0.063	−0.157
	(0.046)	(0.113)	(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 zeros, drops log patent count from the controls, and runs unweighted regressions in order to extend the sample to firms that 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.

are more open to disruption (and willing to hire younger managers) tend to be the ones producing more radical innovations. The quantitative magnitudes are sizable but plausible. For example, a 1-year decrease in CEO age is associated with a 0.171 increase in average citations, which is approximately 1.1 percent of the firm-level weighted mean of our innovation quality variable (15.9).

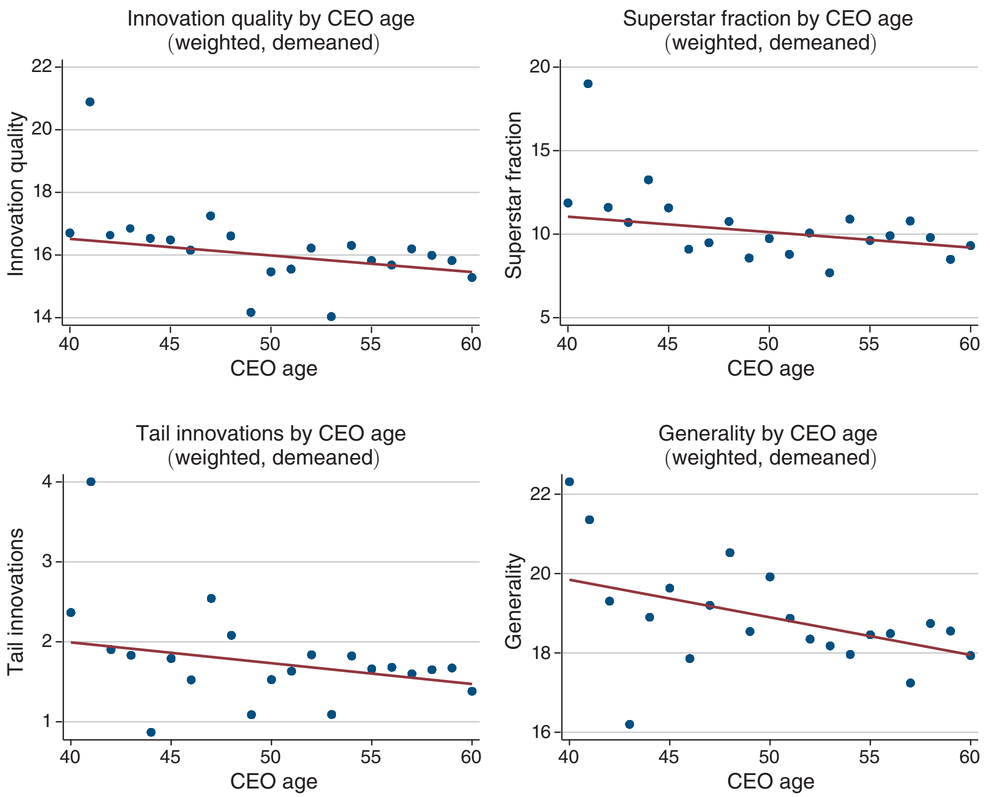


FIGURE 1. THE AVERAGES FOR CREATIVE INNOVATION VARIABLES (INNOVATION QUALITY, SUPERSTAR FRACTION, TAIL INNOVATIONS, AND GENERALITY) BY CEO AGE, AND THE ASSOCIATED FITTED LINE

Notes: The averages are calculated after demeaning the creative innovation variables at the year $\times$ 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.

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 radical innovations are uncorrelated with employment and sales and are only weakly correlated with the number of patents invented 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.[22]

Column 2 of Table 2 shows a similar relationship with the superstar fraction (-0.320, standard error $= 0.132$). This result suggests that younger CEOs tend to work with higher-quality innovators (a relationship we directly investigate in Table 9 below). Columns 3 and 4 show even more precisely estimated relationships with our measures of tail innovations and generality. The implied quantitative magnitudes are

[22] In Table B1 in online Appendix B, we show the same specification as in column 1, but the covariates included one at the time. The results are very stable across columns, which is reassuring.

Table 4—Baseline Panel Regressions

	Innovation quality	Superstar fraction	Tail innovation	Generality
Panel A. Time-varying CEO age				
CEO age	−0.196	−0.260	−0.067	−0.069
	(0.045)	(0.069)	(0.019)	(0.034)
Observations	7,170	7,170	7,170	6,286
Panel B. CEO age (fixed effects)				
CEO age	−0.190	−0.152	−0.049	0.036
	(0.044)	(0.051)	(0.012)	(0.029)
Observations	7,170	7,170	7,170	6,286
Panel C. CEO age and lagged CEO age (fixed effects)				
CEO age	−0.134	−0.103	−0.035	0.031
	(0.041)	(0.039)	(0.013)	(0.026)
Lagged CEO age	−0.125	−0.101	−0.028	0.019
	(0.050)	(0.049)	(0.016)	(0.035)
Observations	5,492	5,492	5,492	4,861
Panel D. CEO age and lagged dependent var (fixed effects)				
CEO age	−0.098	−0.078	−0.037	0.037
	(0.026)	(0.030)	(0.009)	(0.024)
Lagged dependent variable	0.471	0.453	0.214	0.198
	(0.033)	(0.045)	(0.045)	(0.042)
Observations	6,076	6,076	6,076	5,293
Panel E. CEO age and lead CEO age (fixed effects)				
CEO age	−0.115	−0.086	−0.028	0.041
	(0.042)	(0.048)	(0.011)	(0.029)
Lead CEO age	−0.128	−0.114	−0.035	−0.006
	(0.048)	(0.044)	(0.014)	(0.028)
Observations	5,494	5,494	5,494	5,179

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 SIC4 industry fixed effects in panel A and a full set of firm fixed effects in panels B to E (and thus firm age and the four-digit SIC dummies are no longer included). In panels A and B, the key right-hand-side variable is CEO age (in that year). Panel C is identical to panel B except that it also includes a one-year lag of CEO age as well as current CEO age. Panel D is identical to panel B except that it also includes a one-year lag of the dependent variable on the right-hand side. Panel E is identical to panel B except that it also includes a one-year lead of CEO age in addition to current CEO age. See text and notes to Table 1 for variable definitions.

also a little larger—a 1-year increase in CEO age is associated with, respectively, 3.2 percent and 4.6 percent 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 $\times$ year means (and group all observations with CEO age $\leq$ 40 or CEO age $\geq$ 60). 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 online Appendix B.

Table 3 and online Appendix Table B6 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. Perhaps most importantly, 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 analog 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 percent).

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 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.

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

In addition, Table B2 in online Appendix B shows that the results are very similar in a much smaller balanced sample of 297 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 nonpharmaceutical subsample. These subsamples are further studied in Table 5. Table B3 in the online Appendix demonstrates that our results are very similar when we control for patents during the last three years rather than the entire stock of patents. Online Appendix Table B4 confirms that the results are also similar when we explicitly recognize the selection into having any patents using a two-step Heckman correction, while online Appendix Table B5 verifies that these relationships are not driven by firms that have recently undertaken an initial public offering (IPO) by removing firms that have had an IPO over the last ten years from the sample.

Table B6 in the online Appendix shows the weighted and the unweighted relationship between CEO age and several alternative measures of radical 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 radical 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 there are a few cases where the relationship is no longer significant in the unweighted specifications. 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. Consistent with our earlier emphasis, we do not find any relationship between CEO age and R&D intensity. In panels E and F, we show that CEO age also predicts first and second renewals of patents, which are alternative measures of the value and thus creativity of a patent. Also supportive is the fact that CEO age does not predict "internal innovations" (coded from patents that have more than half of their citations to the firm's other patents). Lastly, Table B7 in the online Appendix shows a similar relationship between average CEO age in industry and industry-level measures of radical innovations.

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 radical innovations.

Table 5—Panel Regressions—Robustness

	Innovation quality	Superstar fraction	Tail innovation	Innovation quality	Superstar fraction	Tail innovation
Panel A. Unweighted						
CEO age	−0.182	−0.140	−0.043	−0.181	−0.127	−0.038
	(0.050)	(0.041)	(0.023)	(0.108)	(0.044)	(0.036)
Lead CEO age				−0.022	−0.015	0.010
				(0.125)	(0.058)	(0.043)
Observations	7,170	7,170	7,170	5,472	5,472	5,472
Panel B. Missing observations as zeros						
CEO age	−0.096	−0.085	−0.032	−0.097	−0.067	−0.029
	(0.035)	(0.032)	(0.013)	(0.063)	(0.035)	(0.020)
Lead CEO age				−0.027	−0.030	−0.007
				(0.072)	(0.038)	(0.024)
Observations	11,525	11,525	11,525	10,009	10,009	10,009
Panel C. High-tech subsample						
CEO age	−0.129	−0.089	−0.037	−0.053	−0.031	−0.004
	(0.054)	(0.082)	(0.015)	(0.054)	(0.084)	(0.017)
Lead CEO age				−0.120	−0.103	−0.056
				(0.060)	(0.056)	(0.022)
Observations	2,100	2,100	2,100	1,715	1,715	1,715
Panel D. Low-tech subsample						
CEO age	−0.222	−0.193	−0.056	−0.167	−0.140	−0.047
	(0.068)	(0.064)	(0.017)	(0.067)	(0.058)	(0.014)
Lead CEO age				−0.111	−0.108	−0.020
				(0.068)	(0.056)	(0.018)
Observations	5,070	5,070	5,070	3,757	3,757	3,757
Panel E. Controlling for self-citation fraction						
CEO age	−0.187	−0.151	−0.048	−0.116	−0.090	−0.028
	(0.044)	(0.052)	(0.012)	(0.041)	(0.049)	(0.011)
Lead CEO age				−0.122	−0.110	−0.034
				(0.048)	(0.044)	(0.014)
Observations	6,286	6,286	6,286	5,157	5,157	5,157

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 panels B and E of Table 4 without using weights. Panel B replaces the missing values for dependent variables by zeros, drops log patent count from the controls, and runs unweighted regressions in order to extend the sample to firms that 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 E repeats the same regressions, adding self-citation fraction as a control. See text and notes to Table 1 for variable definitions.

B. *Panel Results with Firm Fixed Effects*

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

Panel A in Table 4 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 radical 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 11 percent larger than the specification without fixed effects in Table 2 (e.g., -0.190 versus -0.171), whereas for superstar fraction and tail innovations, it is smaller—about 48 percent to 62 percent 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 radical 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, 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 implication 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

[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.

[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.

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 have a comparative advantage in radical innovation (for example, because they are more open to disruption), an issue we investigate in greater detail in the next subsection.

Table 5 investigates the robustness of within-firm relationships reported in Table 4, 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 radical 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]

Overall, the results in this subsection demonstrate that firms that switch to younger CEOs generate more radical 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 radical 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.

C. *The Causal Effect of Manager Age on Radical Innovations*

In this subsection, we perform an indirect inference exercise in order to shed further light on the causal effect of manager age on radical innovations. We choose the parameters of the model presented in Section I 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 radical innovations. We

[27] Online Appendix Table B8 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 nonpharmaceutical 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 are reported. Finally, online Appendix Table B9 shows the within-firm equivalent of online Appendix Table B6, where the relationship between CEO age and the alternative measures of radical 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.

perform this exercise both by matching moments from the data and our estimates from the weighted regressions (namely, from panels B and E in Table 4) and from the unweighted regressions (from panel A in Table 5).

The (average) impact of a younger manager on the creativity of innovations for a given firm type is $\frac{1}{F(a^*)}\int_0^{a^*}\Lambda\theta_H\bar{q}^a dF(a) = \frac{\Lambda\theta_H\delta}{g+\delta}\frac{[1-e^{-(g+\delta)a^*}]}{[1-e^{-\delta a^*}]}$. Because of the sorting of younger managers to high-type firms, 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 radical innovations more straightforwardly. For this exercise, we set the discount rate to $\rho = 0.02$ and normalize the 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 $x = 5$ percent, 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 (2018), 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 I and Appendix A). We can then generate simulated firm histories from which the equivalents of the reduced-form regression coefficients in Table 4 can be computed. Of particular importance for this exercise are the specifications in panels B and E of Table 4, where various measures of radical innovations were regressed on current CEO age (and lead CEO age in panel E), 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 B of Table 4 by $\gamma_{current}$ and the coefficient estimates on current and lead CEO age in column 1, panel E, 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 panels B and E of Table 4 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)

Table 6—Structural Parameters

Estimation 1	Estimation 2	Description	Identification
External calibration			
$x = 0.05$	$x = 0.05$	Entry rate	Compustat sample
$\rho = 0.02$	$\rho = 0.02$	Discount rate	Standard value
$\delta = 0.04$	$\delta = 0.04$	Manager death rate	Compustat sample
$\alpha = 0.93$	$\alpha = 0.93$	Reduction rate of innovation size	Akcigit and Kerr (2015)
Indirect inference			
$\psi = 0.079$	$\psi = 0.139$	Baseline radical innovation rate for high type	Estimate
$\Lambda\theta_H = 0.005$	$\Lambda\theta_H = 0.068$	High-type innovation parameter	Estimate
$\varphi = 0.040$	$\varphi = 0.400$	Transition rate from low type to high type	Estimate
$\xi = 0.037$	$\xi = 0.176$	Incremental innovation rate	Estimate
$\eta = 0.433$	$\eta = 0.221$	Initial innovation size	Estimate
$\zeta = 0.272$	$\zeta = 0.910$	Probability of high-type entrant	Estimate

Notes: This table documents the parameter choices and estimates. The first column displays the parameter values obtained in the first estimation, where we target the weighted regression coefficients presented in panel A of Table 8. The second column displays the parameter values obtained in the second estimation, where we target the unweighted regression coefficients presented in panel B of Table 8. See Section IIID for details.

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).[28] This implies that we have in total six data moments and six parameters.

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. 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 6 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. The first column of Table 6 displays the parameter values obtained in the first estimation, where we target the coefficient estimates from the weighted regressions, whereas the second column displays those obtained in the second estimation, where we target the coefficient estimates from the unweighted regressions instead. Table 7 shows the match between the values of these moments in the data and those implied by the model for the two specifications. Panel A in Table 7 reports estimates using weighted regression targets, while panel B is for the case with unweighted regression targets. In both cases, the model-implied numbers are very close to the targeted empirical moments, and

[28] Following Akcigit and Kerr (2018), we define internal patents as those whose majority of citations are self cites.

Table 7—Empirical and Model-Generated Moments

Target	US Data	Model
Panel A. Estimation 1—weighted regression targets		
Current manager age coefficient of Table 4 panel B column 1	−0.190	−0.190
Lead manager age coefficient of Table 4 panel E column 1	−0.128	−0.125
Current manager age coefficient of Table 4 panel E column 1	−0.115	−0.108
Annual growth rate	5.75%	5.96%
Within-firm coefficient of variation of radical innovations	1.99	2.22
Fraction of internal patents	21.5%	21.1%
Panel B. Estimation 2—unweighted regression targets		
Current manager age coefficient of Table 5 panel A column 1	−0.182	−0.187
Lead manager age coefficient of Table 5 panel A column 4	−0.022	−0.022
Current manager age coefficient of Table 5 panel A column 4	−0.181	−0.176
Annual growth rate	5.75%	5.85%
Within-firm coefficient of variation of radical innovations	1.99	1.08
Fraction of internal patents	21.5%	20.7%

Notes: This table displays the empirical and model-generated moments for the indirect inference procedure under the two estimation procedures. In the first estimation, we target the coefficient estimates from the weighted regressions in panels B and E of Table 5. In the second estimation, we target the coefficient estimates from the unweighted regressions in panel A of Table 6. The first three targets are different between the estimations, whereas the last three targets remain the same. See Section IIID for details.

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 nontrivial source of radical innovations.

The implied pattern is also visible in Figures 2 and 3, which plot the probability of a radical innovation and the average CEO age as a function of time since switching to high type for the estimates in panel A of Table 7. 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 radical innovations. Because much of this increase in the probability of radical 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. Using the first set of estimated parameter values, 20 percent 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 percent 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, 3.98 percent and 2.75 percent.

Using the parameter estimates from these exercises, we next compute the "causal effect" of manager age on radical innovations. We start with the equilibrium stationary distribution and then replace them with old managers (in practice, we simply reverse the allocation of managers to firms by age). We assume that after this

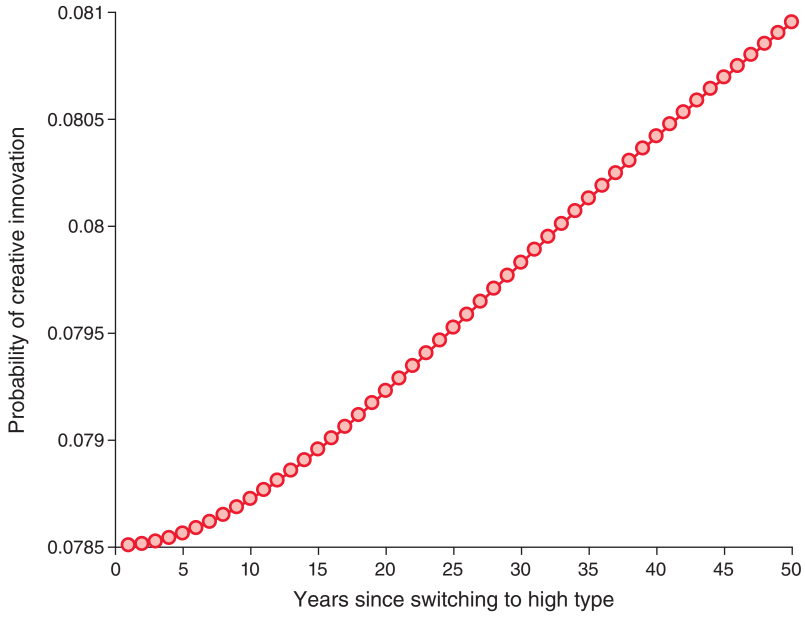


FIGURE 2. EVOLUTION OF CREATIVE INNOVATIONS FOR HIGH-TYPE FIRMS

Notes: 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.

reshuffling, each firm will pursue the same innovation strategy as before.[29] We then calculate the change in radical innovations in this hypothetical economy compared to the baseline economy. This exercise yields a 1 percent decline in the average number of radical innovations with the weighted estimates and 7 percent with the unweighted estimates. Taken together, these results imply that although the causal effect of manager age on radical innovations is positive, it cannot account for the bulk of the variation in radical innovations (which are, instead, mostly driven by firm type according to our estimates).[30]

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 radical innovations is due to sorting. Nonetheless, there is a nonnegligible causal effect of younger managers on radical innovations as well.

[29] 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.

[30] We also note that, even though the causal effect of young managers is small, this is still sufficient in our model to support an allocation in which, as in the data, younger managers are allocated to high-type firms pursuing radical innovations.

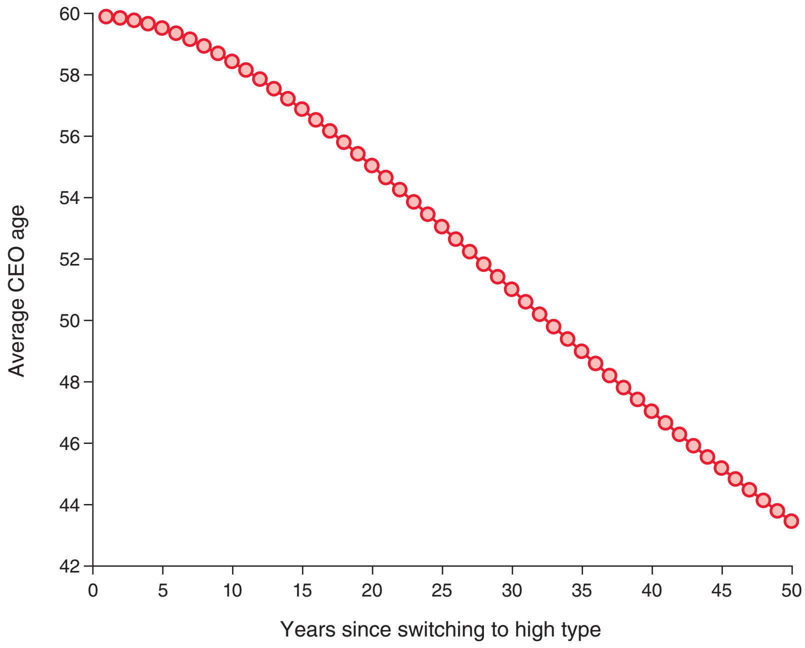


Figure 3. Evolution of CEO Age for High-Type Firms

Notes: 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 0, whereas this number is 26 in the data.

D. *Inventor Age and Creativity of Innovations*

We next turn to an investigation of the role of inventors in radical innovations, focusing again on age. For this purpose, we use patent-level regressions and estimate whether there is an empirical association between inventor age 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 (2014) 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

$$y_{ift} = \phi I_{ift} + \alpha m_{ft} + \mathbf{X}'_{ift}\beta + \delta_f + \gamma_i + d_t + \varepsilon_{ift}. \tag{14}$$

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.[31]

The results from the estimation of (14) are reported in Table 8. In panel A, we focus on a specification similar to the regressions with firm fixed effects reported in Table 4. 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 panel A of Table 4 (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 4, though a little smaller. In column 1, for instance, we see an estimate of -0.121 (standard error $=$ 0.038) compared to -0.190 in Table 4. As a natural patent-level analog 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.[32]

Panel B of Table 8 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,[33] a full set of dummies for the size of the inventor team (i.e., how many inventors are listed), and a full set of dummies for the three-digit IPC class.[34] The inclusion of this

[31] 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 percent of the patents in our sample.

[32] 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 4 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 8.

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

[34] This corresponds to 270 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.

Table 8—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	−0.121	−0.317	−1.241	0.029
	(0.038)	(0.131)	(0.412)	(0.025)
Observations	318,007	318,007	318,007	264,972
Panel B. Inventor age, unbalanced firm sample, 1992–2004				
Inventor age	−0.234	−0.446	−2.873	−0.019
	(0.026)	(0.122)	(0.318)	(0.022)
Observations	318,007	318,007	318,007	264,972
Panel C. Inventor age, extended sample, 1985–2004				
Inventor age	−0.226	−0.380	−2.828	−0.017
	(0.022)	(0.076)	(0.292)	(0.017)
Observations	574,903	574,903	574,903	468,450
Panel D. Inventor age, extended sample, 1985–2004				
Inventor age	−0.199	−0.321	−2.336	−0.046
	(0.010)	(0.035)	(0.131)	(0.011)
Observations	1,879,300	1,879,300	1,879,300	1,560,165
Panel E. CEO age and inventor age, unbalanced firm sample, 1992–2004				
Inventor age	−0.234	−0.444	−2.866	−0.019
	(0.026)	(0.121)	(0.318)	(0.021)
CEO age	−0.121	−0.319	−1.216	0.028
	(0.036)	(0.126)	(0.387)	(0.022)
Observations	318,007	318,007	318,007	264,972

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 ninety-ninth percentile of the citation distribution, dummy for the patent being above the ninetieth 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-hand-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-hand-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.

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.[35] 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

[35] 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.

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 -0.121 (standard error $=$ 0.036) and that on inventor age is -0.234 (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 radical innovations are smaller in Table 8 than those in our baseline firm-level regressions.

Our next results, reported in Table 9, provide some direct evidence on this by looking at the relationship between inventor age and CEO age. We estimate a regression similar to equation (14) 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 9 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 8. The results, which show a positive (even if only marginally significant) relationship, suggest that younger CEOs tend to hire younger inventors, indirectly corroborating the sorting effect emphasized in our theoretical model.[36] Further evidence consistent with this pattern is provided in Tables B10 and B11 in online Appendix B, where we create time-varying measures of the innovation quality of new and existing inventors of the firm (based on the citations of their patents in the past). We then show that both

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

Table 9—Inventor Age and CEO Age, Unbalanced Firm Sample, 1992–2004

	Inventor age (1)	Inventor age (2)
CEO age	0.013	0.013
	(0.006)	(0.006)
Observations	318,007	318,007

Notes: Patent-level panel regressions with annual observations for the unbalanced firm sample 1992–2004. 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.

using our innovation quality measure and the tail innovation index, a younger CEO is associated both with an increase in the creativity of the innovation of continuing inventors and an even larger increase in the quality of new inventors.

E. *Stock of Knowledge*, *Opportunity Cost, and Creativity of Innovations*

Finally, Table 10 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 radical 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 (13). 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 0.69 in our sample), thus stacking the cards against finding an informative set of results.

Nevertheless, Table 10, which uses the same firm sample with annual observations as in panel A of Table 4, 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 for the log patent interaction for the innovation quality measure.[37] These results thus provide some support for the hypothesis that the stock of knowledge of the company and

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

Table 10—Stock of Knowledge, Opportunity Cost, and Creative Innovations, Unbalanced Firm Sample, 1992–2004

	Innovation quality	Superstar fraction	Tail innovation	Generality
CEO age	−0.182	−0.216	−0.057	−0.044
	(0.027)	(0.027)	(0.008)	(0.016)
log sales	1.440	2.053	0.252	1.187
	(0.448)	(0.609)	(0.142)	(0.327)
log patent	−0.389	−0.064	0.085	−0.004
	(0.193)	(0.257)	(0.065)	(0.150)
CEO age × log patent	−0.004	−0.070	−0.016	−0.036
	(0.014)	(0.021)	(0.006)	(0.011)
CEO age × log sales	0.023	0.079	0.018	0.044
	(0.017)	(0.021)	(0.006)	(0.011)
Observations	7,170	7,170	7,170	6,286

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.

opportunity cost effects are present and might in fact be quite important (at least quantitatively at this correlational level).

IV. 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, we undertook a first investigation of the role of firms, managers, and innovators in radical (more creative) innovations. We provided a simple model drawing a clear distinction between radical 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 radical innovation contributions of certain types of firms (for example, those that are more open to disruptions and interested in new technological paradigms), which are more willing to hire younger managers. We do this using several measures of radical innovations, including 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 ninety-ninth 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 or other factors creating a comparative advantage in radical innovations.

We find fairly consistent and robust cross-sectional and within-firm correlations between openness to disruption, proxied by CEO age, and radical innovations. They do not, however, correspond to the causal effect of CEO age on radical innovations because, as highlighted by our theoretical model, younger managers tend to be employed by firms that are "high-type"—for example, 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 radical innovations is due to these sorting effects, and the causal impact of CEO age is positive but small.

Our paper highlights the need for future work investigating the effects of different types of firm organizations and other attributes of managers and innovators. A particularly fruitful direction would be to systematically investigate what types of firms and firm organizations encourage creativity and lead to more radical 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 I

A1. *The Derivation of Equations* (7), (8), *and* (9)

A firm makes the innovation decision in each of its product lines to maximize its present discounted value, which we denote by $W_s(\vec{q_f}, \vec{n_f})$, where $s \in \{H, L\}$, $\vec{q_f}$ is the vector of productivities of the firm, $\vec{n_f}$ 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

$$
\begin{aligned}
\text{(A1)}\quad & rW_L\left(\vec{q_f}, \vec{n_f}\right) - \dot{W}_L\left(\vec{q_f}, \vec{n_f}\right) \\
&= \sum_{m=1}^{m_f} \Bigg[\max_a \left\{ \pi q_{fj_m} + \bar{q}_t f(a) - w_{a,t} \right\} \\
&\qquad + \xi \Big[W_L\Big(\vec{q_f} \backslash \{q_{fj_m}\} \cup \{q_{fj_m} + \eta_{n_{fj_m}+1}\},\ \vec{n_f} \backslash \{n_{fj_m}\} \cup \{n_{fj_m} + 1\} \Big) \\
&\qquad\qquad - W_L\left(\vec{q_f}, \vec{n_f}\right) \Big] \\
&\qquad + \tau \Big[W_L\left(\vec{q_f} \backslash \{q_{fj_m}\}, \vec{n_f} \backslash \{n_{fj_m}\} \right) - W_L\left(\vec{q_f}, \vec{n_f}\right) \Big] \Bigg] \\
&\quad + \varphi \Big[W_H\left(\vec{q_f}, \vec{n_f}\right) - W_L\left(\vec{q_f}, \vec{n_f}\right) \Big].
\end{aligned}
$$

The right-hand side of this value function can be explained as follows: for each product line $m = 1, \ldots, m_f$, the firm receives a revenue stream of πq_{fj_m} as a function of its productivity in this product line, q_{fj_m}. In addition, it has a choice of the age of the manager it will hire to operate this product line (formally choosing $a \in \mathbb{R}_+ \cup \{\varnothing\}$, 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_{fj_m} incremental innovations, the mth element of $\vec{q}_f$ changes from q_{fj_m} to $q_{fj_m} + \eta_{n_{fj_m}+1}$ and n goes up by one. We represent this with the arguments of the value function changing to $\vec{q}_f \backslash \{q_{fj_m}\} \cup \{q_{fj_m} + \eta_{n_{fj_m}+1}\}, \vec{n}_f \backslash \{n_{fj_m}\} \cup \{n_{fj_m} + 1\}$ (and the firm relinquishes its current value function $W_L(\vec{q}_f, \vec{n}_f)$). The firm might also lose one of its currently active product lines to radical destruction, which happens at the endogenous rate τ, and in that case, the firm's value function changes from $W_L(\vec{q}_f, \vec{n}_f)$ to $W_L(\vec{q}_f \backslash \{q_{fj_m}\}, \vec{n}_f \backslash \{n_{fj_m}\})$ (i.e., $\vec{q}_f$ changes $\vec{q}_f \backslash \{q_{fj_m}\}$ and $\vec{n}_f$ to $\vec{n}_f \backslash \{n_{fj_m}\}$). 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, $W_H(\vec{q}_f, \vec{n}_f)$.

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

$$
\begin{aligned}
\text{(A2)}\quad & r W_H(\vec{q}_f, \vec{n}_f) - \dot{W}_H(\vec{q}_f, \vec{n}_f) \\
&= \sum_{m=1}^{m_f} \max\Bigg\{ \pi q_{fj_m} + \max_a \Big\{ \bar{q}_t f(a) - w_{a,t} \\
&\qquad + \xi \Big[W_H\Big(\vec{q}_f \backslash \{q_{fj_m}\} \cup \{q_{fj_m} + \eta_{n_{fj_m}+1}\}, \\
&\qquad\qquad \vec{n}_f \backslash \{n_{fj_m}\} \cup \{n_{fj_m} + 1\}\Big) - W_H(\vec{q}_f, \vec{n}_f) \Big] \Big\}; \\
&\qquad \pi q_{fj_m} + \max_a \Big\{ \bar{q}_t f(a) \\
&\qquad + \Lambda \theta_H \bar{q}^a \Big[E W_H\big(\vec{q}_f \cup \{q_{j'} + \eta_0\}, \vec{n}_f \cup \{0\}\big) \\
&\qquad\qquad - W_H(\vec{q}_f, \vec{n}_f) \Big] - w_{a,t} \Big\} \Bigg\} \\
&\quad + \sum_{m=1}^{m_f} \tau \Big[W_H\big(\vec{q}_f \backslash \{q_{fj_m}\}, \vec{n}_f \backslash \{n_{fj_m}\}\big) - W_H(\vec{q}_f, \vec{n}_f) \Big] \\
&\quad + \sum_{m=1}^{m_f} \psi \Lambda \theta_H \Big[E W_H\big(\vec{q}_f \cup \{q_{j'} + \eta_0\}, \vec{n}_f \cup \{0\}\big) - W_H(\vec{q}_f, \vec{n}_f) \Big].
\end{aligned}
$$

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 (6). In the case of a successful radical innovation, the value of the firm changes to $E W_H(\vec{q}_f \cup \{q_{j'} + \eta_0\}, \vec{n}_f \cup \{0\})$, 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 (7), (8), and (9) follow straightforwardly by conjecturing their forms and verifying this conjecture. ∎

Proof of Propositions 1 and 2.—We present the Proofs of Propositions 1 and 2. The characterization of the rest of the general equilibrium is relegated to online Appendix B.

PROOF OF PROPOSITION 1:

First, substitute the equilibrium wage (10) into (8) to obtain a simplified value function for low-type firms as

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

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: $V_H(q_j,n) = A q_j + \tilde{B}(n)\,\bar{q}_t$. Then the value function of a product line operated by a low-type firm,* (8) *takes the following form*

$$V_L(q_j,n) = A q_j + B(n)\,\bar{q}_t, \tag{A3}$$

where

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

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 (8), we get

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

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

$$rA q_j = \pi q_j - \tau A q_j;$$

and

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

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 (11) into (9) to obtain a simplified form of the value function of a product line operated by a high-type firm as

$$rV_H(q_j,n) - \dot{V}_H(q_j,n)$$
$$= \max\Big\{\pi q_j + \xi\Big[V_H\big(q_j + \bar{q}_t\eta\alpha^{n+1}, n+1\big) - V_H(q_j,n)\Big];$$
$$\pi q_j + \Lambda\theta_H\bar{q}^{a^*}EV_H(\bar{q}_t)\Big\}$$
$$- \tau V_H(q_j,n) + \psi\Lambda\theta_H EV_H(\bar{q}_t).$$

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* (9) *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*

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

where

$$n^* = \min_{n' \in \mathbb{Z}_+} n' \tag{A6}$$

such that

$$\xi\Big[A\eta\alpha^{n'+1} + \tilde{B}\big(n'+1\big) - \tilde{B}\big(n'\big)\Big] \ \leq \ \Lambda\theta_H\bar{q}^{a^*}\Big[\big(1+\eta\big)A + \tilde{B}\big(0\big)\Big].$$

PROOF OF LEMMA 2:

We now conjecture that the value function for high-type firms takes the form in (A4) and substitute this into (9) to obtain

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

Once again equating coefficients, we obtain $A = \pi/(r+\tau)$ and

$$\text{(A7)} \qquad \begin{aligned}(r-g+\tau)\tilde{B}(n) \ = \ & \psi\Lambda\theta_H\Big[A\big(1+\eta\big) + \tilde{B}(0)\Big] \\ & + \max\Big\{\xi\Big[A\eta\alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n)\Big]; \\ & \Lambda\theta_H\bar{q}^{a^*}\Big[(1+\eta)A + \tilde{B}(0)\Big]\Big\}.\end{aligned}$$

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

$$\begin{aligned}(r-g+\tau)\hat{B}(n) \ = \ & \psi\Lambda\theta_H\Big[A\big(1+\eta\big) + \tilde{B}(0)\Big] \\ & + \xi\Big[A\eta\alpha^{n+1} + \hat{B}(n+1) - \hat{B}(n)\Big].\end{aligned}$$

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,

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

where $\beta = \xi/(r-g+\tau+\xi)$ and $\hat{\psi} = \psi\Lambda\theta_H[A(1+\eta) + \tilde{B}(0)]$. 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, Lucas, and Prescott 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, Lucas, and Prescott 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, $\Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$, 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, $\tilde{B}(n) = \frac{\hat{\psi} + \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]}{r-g+\tau}$. Therefore, there exists a smallest integer n^* (which could be zero) such that $\xi[A\eta\alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n)] > \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$ for all $n < n^*$, which verifies the definition of n^* in (A6). By the definition of n^*, we have that $\xi[A\eta\alpha^{n^*+1} + \tilde{B}(n^*+1) - \tilde{B}(n^*)] \leq \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$. Now there are two cases to consider:

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

(2) $\xi[A\eta\alpha^{n^*+1} + \tilde{B}(n^*+1) - \tilde{B}(n^*)] = \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$. 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

$$\text{(A9)} \qquad \lambda_{n^*} \leq 1, \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A + \tilde{B}(0)\right] - \xi\left[A\eta\alpha^{n^*+1} + \tilde{B}\left(n^*+1\right) - \tilde{B}\left(n^*\right)\right] \geq 0$$

and

$$\left[\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A + \tilde{B}(0)\right] - \xi\left[A\eta\alpha^{n^*+1} + \tilde{B}\left(n^*+1\right) - \tilde{B}\left(n^*\right)\right]\right] \times \left(1 - \lambda_{n^*}\right) = 0.$$

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 that 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

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

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, Lucas, and Prescott 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

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

and thus

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

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 $\xi\tilde{b}(n+1) + \xi c \leq \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$ (case (i)). In this case,

$$\text{(A12)} \quad T\big(\tilde{b}(n+1) + c\big) = F^{-1}\Big[(r - g + \tau)A\eta\alpha^n + \Lambda\theta_H\bar{q}^{a^*}\big[(1+\eta)A + \tilde{B}(0)\big]\Big]$$
$$= T\big(\tilde{b}(n+1)\big).$$

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

$$\text{(A13)} \quad T\big(\tilde{b}(n) + c\big) = F^{-1}\Big[(r - g + \tau)A\eta\alpha^n + \xi\tilde{b}(n+1) + \xi c\Big]$$
$$= \frac{r - g + \tau}{r - g + \tau + \xi}A\eta\alpha^n + \frac{\xi}{r - g + \tau + \xi}\tilde{b}(n+1)$$
$$+ \frac{\xi}{r - g + \tau + \xi}c$$
$$\leq T\big(\tilde{b}(n+1)\big) + \frac{\xi}{r - g + \tau + \xi}c,$$

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 fourth line follows because

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

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

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

Then, by setting $\beta = \xi/(r - g + \tau + \xi) < 1$, 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^*, $\tilde{b}(n^*) \leq \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$, and since $\tilde{b}$ is strictly decreasing, we have $\tilde{b}(n^*+1) < \Lambda\theta_H\bar{q}^{a^*}[(1+\eta)A + \tilde{B}(0)]$. 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 Lemmas 1 and 2. ∎

The derivation of the stationary distribution and proof of existence of general equilibrium are provided in online Appendix B.

PROOF OF PROPOSITION 2:

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

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

and

$$\begin{aligned} & rV_H(q_j,n) - \dot{V}_H(q_j,n) \\ & \quad = \max\Big\{\pi q_j + \max_a\Big\{\bar{q}_t f(a) - w_{a,t} + \xi\big[V_H(q_j + \eta_{n+1}, n+1) - V_H(q_j,n)\big]\Big\}; \\ & \qquad\qquad \pi q_j + \max_{a \geq 0}\{\bar{q}_t f(a) + \Lambda\theta_H\bar{q}^a EV_H(t) - w_{a,t}\}\Big\} \\ & \qquad - \tau V_H(q_j,n) + \psi\Lambda\theta_H EV_H(t). \end{aligned}$$

Here note that, with a slight abuse of notation, we wrote $EV_H(t)$ instead of $EV_H(\bar{q}_t)$ 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

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

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

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

and

$$\begin{aligned} r\,V_H(q_j,n) - \dot{V}_H(q_j,n) \;=\;& \max\Big\{\pi\,q_j + \xi\big[V_H(q_j+\eta_{n+1},n+1) - V_H(q_j,n)\big]; \\ & \qquad \pi\,q_j + \Lambda\,\theta_H\bar{q}^{a^*}E\,V_H(t)\Big\} \\ & - \tau\,V_H(q_j,n) + \psi\Lambda\,\theta_H E\,V_H(t). \end{aligned}$$

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

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

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,

$$\text{(A14)} \qquad V_H\big(q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*),n+1\big) - V_H(q_{n,t}^*,n) \;=\; \frac{\Lambda\,\theta_H\bar{q}^a}{\xi}E\,V_H(t),$$

and at this value, we also have

$$\text{(A15)} \qquad \begin{aligned} (r+\tau)\,V_H(q_{n,t}^*,n) - \dot{V}_H(q_{n,t}^*,n) \;=\;& \pi\,q_{n,t}^* + \Lambda\,\theta_H\bar{q}^a E\,V_H(t) \\ & + \psi\Lambda\,\theta_H E\,V_H(t). \end{aligned}$$

Now we will consider two alternative cases:

Case 1:

$$q^*_{n+1,t} \;\geq\; q^*_{n,t} + \eta_{n+1,t}\big(q^*_{n,t}\big). \tag{A16}$$

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

$$\begin{aligned}(r+\tau)\,V_H\big(q^*_{n,t} + \eta_{n+1,t}(q^*_{n,t}), n+1\big) - \dot{V}_H\big(q^*_{n,t} + \eta_{n+1,t}(q^*_{n,t}), n+1\big) \\ = \pi q^*_{n,t} + \pi\eta_{n+1,t}\big(q^*_{n,t}\big) + \Lambda\theta_H\bar{q}^a E V_H(t) + \psi\Lambda\theta_H E V_H(t).\end{aligned} \tag{A17}$$

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 (A14) with respect to time, we have

$$\begin{aligned}\dot{V}_H\big(q^*_{n,t} + \eta_{n+1,t}(q^*_{n,t}), n+1\big) - \dot{V}_H\big(q^*_{n,t}, n\big) &= \frac{\Lambda\theta_H\bar{q}^a}{\xi}\,\partial E V_H(t)/\partial t \\ &= \frac{\Lambda\theta_H\bar{q}^a}{\xi}\,gEV_H(t),\end{aligned} \tag{A18}$$

where, in the second line, we have used the fact that in a stationary equilibrium $EV_H(t)$ grows at the rate g. Subtracting (A15) from (A17) and using (A18), we obtain

$$\begin{aligned}(r+\tau)\Big[V_H\big(q^*_{n,t} + \eta_{n+1,t}(q^*_{n,t}), n+1\big) - V_H\big(q^*_{n,t}, n\big)\Big] \\ = \pi\eta_{n+1,t}\big(q^*_{n,t}\big) + \frac{\Lambda\theta_H\bar{q}^a}{\xi}\,gEV_H(t).\end{aligned} \tag{A19}$$

Then, combining (A14) and (A19), we can derive

$$\pi\eta_{n+1,t}\big(q^*_{n,t}\big) \;=\; \frac{r-g+\tau}{\xi}\,\Lambda\theta_H\bar{q}^a E V_H(t). \tag{A20}$$

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

Now let us define

$$v_t \;\equiv\; \frac{r-g+\tau}{\pi\xi}\,\Lambda\theta_H\bar{q}^a E V_H(t), \tag{A21}$$

which is independent of both q and n. Using (A21), equation (A20) can be written as

$$\big[\kappa\bar{q}_t + (1-\kappa)\,q^*_{n,t}\big]\eta\alpha^{n+1} \;=\; v_t, \tag{A22}$$

or

$$q^*_{n,t} \;=\; \frac{v_t/\eta\alpha^{n+1} - \kappa\bar{q}_t}{1-\kappa}. \tag{A23}$$

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 (A16) indeed applies. For this reason, note that from (A22) written for $n + 2$ incremental innovations, we have

$$q^*_{n+1,t} = \frac{v_t/\eta\alpha^{n+2} - \kappa\bar{q}_t}{1-\kappa}. \tag{A24}$$

Combining equations (A23) and (A24), we obtain that (A16) is satisfied if

$$(1-\kappa)\eta\alpha^{n+2} + \alpha \leq 1. \tag{A25}$$

Thus, whenever (A25) 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 (A25) holds, the same result applies.

Case 2:

$$q^*_{n+1,t} - \eta_{n+1,t}(q^*_{n,t}) < q^*_{n,t}. \tag{A26}$$

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 nth prior incremental innovations. This condition is clearly the complement of (A16).

In this case, start with $q^*_{n+1,t}$, which satisfies (A17). Under condition (A26), $q^*_{n,t}$ satisfies (A15), so we again arrive at (A14), (A20), and (A23). But then from (A23) $q^*_{n,t}$ 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 (A25) holds. Note that the same expressions for $q^*_{n+1,t}$ as in (A24) again apply under Case 2. Thus, the condition for (A26) to be satisfied, with an identical argument, is

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

which is indeed the complement of (A25).

Consequently, regardless of whether (A25) or its converse holds, equation (A23) applies, and $q^*_{n,t}$ is increasing in n (or equivalently, $n^*_t(q)$ is increasing in q). This completes the proof. ∎

A2. *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. 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.

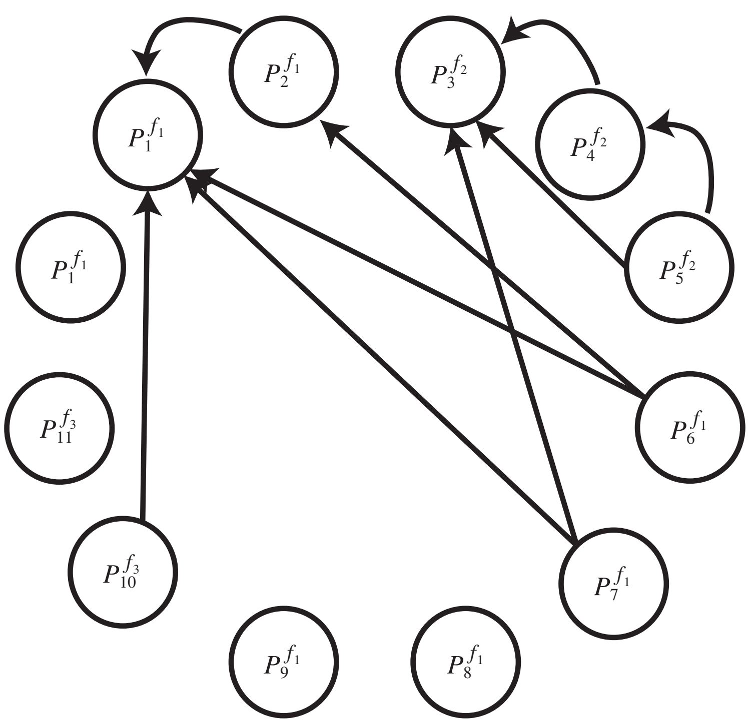
Example 1: The following chart provides an illustrative example focusing on two product lines:

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

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

Here $P_n^{f_m}$ denotes patent n belonging to firm f_m, and η_n denotes its step size as described in equation (5). 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 above, $P_7^{f_1}$ 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.

REFERENCES

Abrams, David S., Ufuk Akcigit, and Jillian Grennan. 2018. "Patent Value and Citations: Creative Destruction or Strategic Disruption?" NBER Working Paper 19647.

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. 2018. "Innovation, Reallocation and Growth." *American Economic Review* 108 (11): 3450–91.

Acemoglu, Daron, Ufuk Akcigit, and Murat A. Celik. 2022. "Replication data for: Radical and Incremental Innovation: The Roles of Firms, Managers, and Innovators." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E125161V1.

Akcigit, Ufuk, and William R. Kerr. 2018. "Growth through Heterogeneous Innovations." *Journal of Political Economy* 126 (4): 1374–1443.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2): 277–97.

Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Incentive Activity: Economic and Social Factors*, 609–26. Princeton, NJ: Princeton University Press.

Azoulay, Pierre, Joshua S. Graff Zivin, and Gustavo Manso. 2011. "Incentives and Creativity: Evidence from the Academic Life Sciences." *RAND Journal of Economics* 42 (3): 527–54.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics* 125 (2): 549–89.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun. 2020. "CEO Behavior and Firm Performance." *Journal of Political Economy* 128 (4): 1325–69.

Barker, Vincent L., III, and George C. Mueller. 2002. "CEO Characteristics and Firm R&D Spending." *Management Science* 48 (6): 782–801.

Benmelech, Efraim, and Carola Frydman. 2015. "Military CEOs." *Journal of Financial Economics* 117 (1): 43–59.

Bennedsen, Morten, Francisco Pérez-González, and Daniel Wolfenzon. 2020. "Do CEOs Matter? Evidence from Hospitalization Events." *Journal of Finance* 75 (4): 1877–1911.

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *Quarterly Journal of Economics* 118 (4): 1169–1208.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122 (4): 1351–1408.

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–24.

Chari, V.V., and Hugo Hopenhayn. 1991. "Vintage Human Capital, Growth, and the Diffusion of New Technology." *Journal of Political Economy* 99 (6): 1142–65.

Christensen, Clayton M. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.* Boston, MA: Harvard Business Review Press.

Fogli, Alessandra, and Laura Veldkamp. 2021. "Germs, Social Networks and Growth." *Review of Economic Studies* 88 (3): 1074–1100.

Foster, Lucia, John C. Haltiwanger, and C.J. Krizan. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." In *New Developments in Productivity Analysis,* edited by Charles R. Hulten, Edwin R. Dean, and Michael J. Harper, 303–72. Chicago: University of Chicago Press.

Galenson, David W., and Bruce A. Weinberg. 2000. "Age and the Quality of Work: The Case of Modern American Painters." *Journal of Political Economy* 108 (4): 761–77.

Galenson, David W., and Bruce A. Weinberg. 2001. "Creating Modern Art: The Changing Careers of Painters in France from Impressionism to Cubism." *American Economic Review* 91 (4): 1063–71.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5): 874–904.

Garicano, Luis, and Esteban Rossi-Hansberg. 2004. "Inequality and the Organization of Knowledge." *American Economic Review* 94 (2): 197–202.

Gorodnichenko, Yuriy, and Gerard Roland. 2017. "Culture, Institutions and the Wealth of Nations." *Review of Economics and Statistics* 99 (3): 402–16.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

Harhoff, Dietmar, Francis Narin, F.M. Scherer, and Katrin Vopel. 1999. "Citation Frequency and the Value of Patented Inventions." *Review of Economics and Statistics* 81 (3): 511–15.

Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. 2nd ed. Thousand Oaks, CA: Sage.

Hurst, Erik, and Benjamin Wild Pugsley. 2011. "What Do Small Businesses Do?" *Brookings Papers on Economic Activity* 42 (2): 73–142.

Jones, Benjamin 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, Benjamin F. 2010. "Age and Great Invention." *Review of Economics and Statistics* 92 (1): 1–14.

Jones, Benjamin F., and Bruce A. Weinberg. 2011. "Age Dynamics in Scientific Creativity." *Proceedings of the National Academy of Sciences* 108 (47): 18910–14.

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

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological Innovation, Resource Allocation, and Growth." *Quarterly Journal of Economics* 132 (2): 665–712.

Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming. 2014. "Disambiguation and Coauthor Networks of the US Patent Inventor Database." *Research Policy* 43 (6): 941–55.

Lentz, Rasmus, and Dale T. Mortensen. 2008. "An Empirical Model of Growth through Product Innovation." *Econometrica* 76 (6): 1317–73.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (2): 508–23.

MacDonald, Glenn, and Michael S. Weisbach. 2004. "The Economics of Has-Beens." *Journal of Political Economy* 112 (S1): S289–310.

Sampat, Bhaven N. and Arvids A. Ziedonis. 2005. "Patent Citations and the Economic Value of Patents." In *Handbook of Quantitative Science and Technology Research*, edited by Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch, 277–98. Berlin: Springer.

Sarada, and Tocoian, O. 2019. "Is It All About Who You Know? Prior Work Connections and Entrepreneurial Success." *ILR Review* 72 (5): 1200–1224.

Schumpeter, Joseph A. 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.

Standard and Poor's. 2012. "Compustat North America Fundamentals Annual." Wharton Research Data Services. https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/sp-global-market-intelligence/ (accessed October 2012).

Stokey, Nancy L., Robert E. Lucas, Jr., and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

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

Trajtenberg, Manuel. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *RAND Journal of Economics* 21 (1): 172–87.

Weinberg, Bruce A., and David W. Galenson. 2019. "Creative Careers: The Life Cycles of Nobel Laureates in Economics." *De Economist* 167: 221–39.

Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics* 113 (2): 331–60.