Does the Cream Always Rise to the Top? The Misallocation of Talent in Innovation*

Murat Alp Celik†

November 11, 2015

Please click here to download the most recent version.

Abstract

The misallocation of talent between routine production versus innovation activities is shown to have a first-order impact on the welfare and growth prospects of an economy. Surname level empirical analysis combining patent and inventor micro-data with census data reveals new stylized facts: (1) People from richer backgrounds are more likely to become inventors; but those from more educated backgrounds are not. (2) People from more educated backgrounds become more prolific inventors; but those from richer backgrounds exhibit no such aptitude. Motivated by this discrepancy, a heterogeneous agents model with production and innovation sectors is developed. Individuals compete against each other for scarce inventor training in a tournament setting. Those from richer families can become inventors even if they are of mediocre talent by excessive spending on credentialing. This is individually rational but socially inefficient. The model is calibrated to match the new stylized facts via indirect inference. A thought experiment in which the credentialing spending channel is shut down reveals that the rate of innovation can be increased by 10% of its value. Optimal progressive bequest taxes serve to increase social welfare by 6.20% in consumption equivalent terms.

Keywords: Economic Growth, Inequality, Innovation, Misallocation, Patents, Optimal Taxation

JEL Classification: O15, O31, O41

*I would like to thank my advisors Ufuk Akcigit and Jeremy Greenwood; as well as Daron Acemoglu, Philippe Aghion, Paul Beaudry, Gregory Clark, Harold Cole, Flavio Cunha, Selman Erol, Nir Jaimovich, Dirk Krueger, Iourii Manovskii, Daniel Neuhann, Claudia Olivetti, Guillermo Ordonez, Jesse Perla, Michael Peters, Jose-Victor Rios-Rull, Nicolas Roys, Peter Rupert, Ananth Seshadri, Gustavo Ventura, Fabrizio Zilibotti, and the workshop and conference participants at Bilkent University, FRB Philadelphia, University of Nottingham and University of Pennsylvania for helpful comments and discussions. I gratefully acknowledge financial support from the Maloof Family Dissertation Fellowship in Economics and SAS Dissertation Completion Fellowship.

†University of Pennsylvania, Department of Economics, 3718 Locust Walk, Philadelphia, PA, 19104. e-mail: celik@sas.upenn.edu.

1 Introduction

Albert Einstein was born in Ulm on March 14, 1879. His father was Hermann Einstein, a rich salesman and engineer, and owned a company called Elektrotechnische Fabrik J. Einstein & Cie that manufactured electrical equipment based on direct current. Albert received his education in various high quality schools in Germany, Italy, and Switzerland, and his alma mater was ETH Zurich. As a scientist and inventor, he produced over 300 scientific papers and 50 patented inventions. His groundbreaking contributions in the field of physics changed the technological landscape. What would happen, though, if his parents were poor and he could not receive the education he had? How would a world look like with Einstein as a factory worker instead of a scientist? Better yet, how do we know if we are not missing out on potential Einsteins right now?

Allocation of talent—assigning the right people to the right jobs—can have a first-order effect on the productivity of a society. The susceptibility of the allocation mechanism to be distorted away from the socially optimal outcome by private expenditures might create significant welfare losses in the presence of high levels of inequality in private resources. The losses are especially magnified if the best and the brightest of a society are not allocated to the professions where their social contributions would be the greatest. This paper aims to quantify the misallocation of talent in the United States due to economic inequality, with particular emphasis on its effects on innovation, and hence the long-run prospects of the country.

Parents spend considerable time and resources in order to improve the likelihood that their children end up with a desired job. The education system serves two main purposes in this regard: improving human capital, and credentialing people's talents.[1] The credentialing part can be seen as a tournament in which individuals seek to improve their overall ranking compared to others in order to improve their job market prospects. In 2010, the United States spent 7.3% of its gross domestic product on education, and the share of private spending was 7.7% for non-tertiary and 63.7% for tertiary education (OECD, 2013). Annual expenditure per student in tertiary education was $25,575, with total yearly cost going up to $60,000 for elite universities. At the same time, the net wage of the median worker was $26,364, whereas the median inventor earned above $100,000 per year. In such a high stakes environment where both the rewards and means to achieve them are unequal, financial frictions can easily prevent the talented children of poor families from being assigned to jobs suitable to their abilities since they are crowded out by the less talented children from richer families. Is this actually what happens in reality, or can we conclude that "the cream always rises to the top" regardless of inefficiencies of the system?

The first contribution of this paper is to provide empirical evidence on the misallocation of talent

[1] Work on education's role in improving human capital is discussed in detail in the related literature subsection. For the use of education in credentialing people's talents (as a "signaling" device), see Spence (1973), Stiglitz (1975) and Fernandez and Gali (1999).

in innovation. Information on innovation in the United States is obtained from inventor and patent level data from United States Patent and Trademark Office (USPTO). This data includes all patents granted in the United States between 1976 and 2006, as well as all registered inventors of these innovations. Inventors are identified uniquely throughout their careers, but direct information on their parental backgrounds is unavailable. In order to overcome this issue, surnames of the inventors are used as a proxy, and the inventor data is linked to socioeconomic background information at the surname level from U.S. census data (1930). The stylized facts obtained can be summarized as follows:[2]

Fact 1: Individuals from richer backgrounds are much more likely to become inventors (23.9%); whereas those from more educated backgrounds experience no similar advantage (0.1%).

Fact 2: Conditional on becoming an inventor, individuals from more educated backgrounds turn out to be much more prolific inventors (17.5%); whereas those from richer backgrounds exhibit no such aptitude (0.1%).

When the two facts are considered together, it appears that the misallocation of talent is an issue for inventors. Fact 2 shows that it is the education associated with the surname and not income that predicts higher inventor quality today. This is intuitive, since education and (unobserved) innate ability are likely to be complementary (or at least highly correlated), and in the presence of persistence of innate ability across generations, one would expect the descendants of the more educated to be better inventors today conditional on becoming one. However, Fact 1 shows that it is income and not education that predicts higher chances of becoming an inventor today. This can be interpreted as the allocation system choosing the wrong people as inventors. Those who come from families that were wealthier but had average education in the past have a higher chance of becoming inventors, but perform poorly conditional on becoming one. This observed discrepancy provides the motivation to investigate the issue of misallocation of talent in innovation quantitatively, so that its impact on the society can be assessed.[3]

In order to quantify the effects of the misallocation of talent in innovation and to analyze potential policy changes that might alleviate the inefficiency, a new model which can accommodate the observed correlation patterns is developed. The firm side exhibits features found in the models from

[2]The numbers in parentheses correspond to how much one standard deviation increase in the independent variable causes the dependent variable to increase compared to its own standard deviation. The details of the empirical analysis can be found in Section 3.

[3]It is also noteworthy that the family background measures have such a high explanatory power. For instance, it is found that one standard deviation increase in the income associated with the surname in 1930 increases the relative probability of becoming an inventor by 23.9%. Given that the measures are constructed at the surname-level, and across two to three generations, just knowing the surname of an individual makes it possible to predict his or her chances of becoming an inventor to a high degree. This means the intergenerational mobility in socioeconomic status as captured by the relative probability of becoming an inventor is quite low, which is consistent with other recent name and surname level studies [Clark (2014), Olivetti and Paserman (2013)].

the endogenous growth literature: Firms undertake routine production using unskilled labor, and generate productivity-improving innovations (featuring positive intertemporal spillovers between firms) via research and development conducted by hired inventors. The household side is modeled in a detailed fashion, borrowing from heterogeneous agents models[4] in order to make the model capable of replicating the patterns observed in the data. The households are heterogeneous in wealth, education, and unobserved innate ability that is persistent across generations. Parents invest in the education of their offspring and leave bequests. The training necessary to become inventors is scarce; hence individuals compete against each other in a tournament setting to receive it. Factors that improve inventor productivity such as innate ability and education increase the probability of receiving this training; but so does private credentialing spending which is unproductive by itself. Thus, individuals who inherit generous bequests can become inventors even if they are of mediocre talent through excessive spending on credentialing, preventing more talented individuals from poorer backgrounds from becoming one. This is individually rational but socially inefficient; reducing the quality of the inventor pool used in generating productivity-improving innovations that drive economic growth.

The tournament mechanism is the key ingredient that enables the model to replicate the stylized facts. In an ideal world, a social planner would prefer to allocate the best and the brightest of the society to the innovation sector, leading to a positive assortative matching between the talents of individuals and the (social as well as private) productivity of the jobs. However, if this were the case, the discrepancy between the parental backgrounds of those who become inventors and those who succeed as inventors would not be empirically observed. In order to allow the model to generate different correlation patterns at the two margins, individuals receive inventor training based on a score that depends differentially on innate ability, early childhood education and credentialing spending. The strength of each component in improving inventor probability as opposed to inventor productivity has different implications for the correlations of ancestor education and income with the two outcome variables, and this provides the main identification in the calibration of the model.

The model is calibrated to match the new stylized facts and data moments from the U.S. economy where an exercise in indirect inference pins down the influence of the new credentialing spending channel by replicating the two regressions from the empirical analysis using model-generated data. The calibrated model is then used to measure the economic importance of the misallocation of talent in innovation. A thought experiment in which the credentialing spending channel is shut down reveals that the aggregate growth rate of the economy can be increased by 10% of its value by assigning more talented and better educated individuals as inventors. As a result, the consumption inequality in the economy increases, which is detrimental to overall welfare; however the gain in

[4] This is in the spirit of Aiyagari (1994) since the heterogeneity of households is considered in a general equilibrium setting, where the time-invariant distributions of household characteristics affect the prices and the growth rate in the economy.

output growth rate more than compensates for this loss, resulting in a welfare gain of 5.96% in consumption equivalent terms.[5]

Seeking to alleviate the effects of misallocation in a decentralized economy, optimal progressive bequest taxes are calculated, which are found to increase output growth rate by 2.5% of its value. This increase is again through the allocation of higher innate ability individuals as inventors, who are also more educated on average. The progressive nature of the taxes cause the overall consumption inequality to remain the same. The increase in the output growth rate and the relatively unchanged consumption inequality lead to a social welfare gain of 6.20% in consumption equivalent terms. This is higher compared to the credentialing spending shut-down experiment. The optimal bequest tax policy that achieves these results is quite progressive: The average bequest tax rate faced by the top 1% is 12.1%, whereas this number falls to 4.2% for the top 10%. The bottom 95% of the households are net recipients, whereas only the top 5% pay into the system.

Related literature: The paper relates to the growing literature on misallocation (some examples are Acemoglu, Akcigit, Bloom, and Kerr (2013), Akcigit, Celik, and Greenwood (2015), Guner, Ventura, and Xu (2008), Hsieh and Klenow (2009), Hsieh, Hurst, Jones, and Klenow (2013), Jones (2013), Jovanovic (2014), and Restuccia and Rogerson (2008)). One of the closest papers in this literature is Hsieh, Hurst, Jones, and Klenow (2013) where the misallocation of talent results from barriers to entry into certain occupations by distinct demographic groups, like women or non-white individuals. Another close paper is Jovanovic (2014) where workers and jobs are heterogeneous in quality, and are matched with each other under search frictions which affects the amount of on-the-job training, and the transition to the balanced growth path. This paper differs from both works by its emphasis on the financial frictions channel, and the special interest on how innovation activities are influenced as a result. Empirically, the closest study is the ongoing work by Bell, Chetty, Jaravel, Petkova, and van Reenen (2015) which focuses on the life cycle of inventors using administrative data covering the population of patent applicants in the U.S. Their finding that individuals from higher income family backgrounds are more likely to become inventors is in agreement with the findings of this paper. This paper provides further evidence on the differential effects of ancestor income and education on becoming an inventor and success conditional on becoming one, as well as build a quantitative model to assess the economic costs of the misallocation of talent in innovation.

Another closely related field is the modern literature on inequality and economic development (see among others: Galor and Zeira (1988, 1994), Banerjee and Newman (1993), Maoz and Moav (1999), and Galor (2009) for a literature review). This paper differs from the literature in that it acknowledges the scarce nature of training necessary to become inventors, and focuses on how

[5] Welfare is defined as expected utility at the steady state.

competition for this might create a misallocation of talent between routine production and innovation. In addition, economic growth in this model is driven by technological change as a result of firms investing in innovative activities, similar to the literature on endogenous growth with quality improvements pioneered by Aghion and Howitt (1992), and in the spirit of the broader endogenous growth literature [Lucas (1988), Romer (1990), Lucas (2009), Alvarez, Buera, and Lucas (2013), Lucas and Moll (2014). See Aghion and Howitt (2009), Acemoglu (2009) and Aghion, Akcigit, and Howitt (2013) for literature surveys]. The firm side of the model builds upon Akcigit, Celik, and Greenwood (2015). Recent work by Aghion, Akcigit, Bergeaud, Blundell, and Hemous (2015) investigates the relationship between innovation and top income inequality. The quantitative results of the current paper are in line with their empirical finding of a positive correlation between the two.

The focus on who become inventors versus who make prolific inventors conditional on becoming one links this work to the extensive literature on nature versus nurture, human capital and skill formation [Becker (1964), Ben-Porath (1967), Behrman, Taubman, and Wales (1977), Becker and Tomes (1979), Becker and Tomes (1986), Behrman, Rosenzweig, and Taubman (1994), Aiyagari, Greenwood, and Seshadri (2002), Heckman, Stixrud, and Urzua (2006), Cunha and Heckman (2007), Dahl and Lochner (2012), Lee and Seshadri (2014). See Cunha, Heckman, Lochner, and Masterov (2006) for a survey]. This literature is quite diverse, ranging from theoretical work such as the classic Becker and Tomes (1979) model, to empirical estimates exploiting rare datasets such as that on twins (Behrman, Rosenzweig, and Taubman, 1994) to separate the effects of nature and nurture. This paper investigates a related question, but focuses on inventors and their productivities in coming up with disruptive inventions as captured by patents. This enables the use of the two new stylized facts obtained in the empirical analysis to tease out the persistence of innate ability versus the socioeconomic status persistence due to intergenerational wealth transmission. The model is close in spirit to Becker and Tomes (1979) type models, where parents cannot borrow against the future income of their dynasties, or insure themselves against idiosyncratic risks.

Finally, the policy experiment on optimal taxation of bequests links the paper to the literature on optimal taxation [Anderberg (2009), Bohacek and Kapicka (2009), Findeisen and Sachs (2013), Grochulski and Piskorski (2010), Kapicka (2013), Kapicka and Neira (2013), Krueger and Ludwig (2013), Stantcheva (2014)]. Two close papers in this field are Krueger and Ludwig (2013) and Stantcheva (2014), where optimal progressive taxation and education subsidies are calculated in a model with heterogeneous households where human capital formation is also endogenous. The model in this paper also includes the endogenous human capital aspect, but enhances the problem by adding in the misallocation of talent dimension and its effects on innovation. This naturally affects the effectiveness of different policies in alleviating the inefficiencies that result from financial frictions.

Outline: The rest of the paper is organized as follows: Section 2 presents the theoretical model. Section 3 describes the datasets employed and variables constructed in the empirical analysis, and the resulting stylized facts. Section 4 describes the calibration of the model and the indirect inference. Section 5 presents and discusses the results of the quantitative results. Section 6 concludes.

2 Model

2.1 Environment and preferences

Time is discrete, and denoted by $t = 0, 1, 2, ...$ There is a continuum of households indexed by $m \in [0, 1]$. The households are modeled in an overlapping generations framework, where each generation lives for three periods: child, young adult and old adult. The children are born when their parents are young adults. The parents interact with their children in three ways: Parents (i) choose their consumption before they become adults, (ii) invest in their education[6] and (iii) leave non-negative bequests to them upon death. The parents care about their children, and the relative weight of the utility of their offspring is denoted by the altruism parameter $\alpha > 0$. Preferences over consumption are time separable with time discount factor β and exhibit constant relative risk aversion with parameter ω. Thus, lifetime utility of generation born at time t of household m can be expressed as

$$U_{m,t}(\vec{c}_{m,t}) = E_t \left[\frac{c_{c,m,t}^{1-\omega}}{1-\omega} + \beta \frac{c_{y,m,t}^{1-\omega}}{1-\omega} + \beta^2 \frac{c_{o,m,t}^{1-\omega}}{1-\omega} + \alpha\beta U_{m,t+1}(\vec{c}_{m,t+1}) \right] \tag{1}$$

where $c_{c,m,t}, c_{y,m,t}$ and $c_{o,m,t}$ denote the consumption of generation t of household m at child, young, and old periods respectively, and $\vec{c}_{m,t} = \{c_{c,m,T}, c_{y,m,T}, c_{o,m,T}\}_{T=t}^{\infty}$.

2.2 Technology

2.2.1 Production and innovation

The final good is competitively produced by a continuum of firms indexed by $i \in [0, 1]$ which combine capital k and unskilled labor l_u according to the formula

$$o(z, k, l_u) = z^{\zeta} k^{\kappa} l_u^{\lambda} \tag{2}$$

where z stands for the firm-specific productivity, o denotes final good output, and $\zeta + \kappa + \lambda = 1$. Firms pay $r + \delta$ and w_u for capital and unskilled labor services respectively.

[6] The education investment of the parents is thought of as pre-tertiary education or early childhood investment as discussed in Cunha and Heckman (2007). Individuals invest in their own tertiary education when they become young adults, which is discussed later on.

Firms can engage in risky innovation activities in order to increase their productivity if successful. Conditional on successful innovation, the productivity of the firm in the next period is evolves according to

$$z' = z + \gamma \bar{z} \tag{3}$$

where z' and z are the new and old productivities, $\bar{z}$ is the average productivity in the economy, and $\gamma > 0$ is a scale parameter.[7] The firms which fail to innovate retain their old productivity, $z' = z$. In order to increase the probability of successful innovation, firms must hire skilled labor. For a firm which hires l_s amount of skilled labor, the probability of a successful innovation is given by

$$i(l_s) = \chi l_s^{\xi} \tag{4}$$

where $\chi > 0$ is a scale parameter and $\xi \in (0, 1)$ introduces diminishing returns.

2.2.2 Individual productivity, innate ability and early childhood education

Each generation t of each household m is heterogeneous in innate ability a, and early childhood education h. The individual productivity of generation t of household m is a constant elasticity of substitution (CES) aggregate of a and h given by

$$l_{m,t}(h_{m,t}, a_{m,t}) = \left(\psi h_{m,t}^{\frac{\epsilon-1}{\epsilon}} + (1-\psi) a_{m,t}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{5}$$

where $0 < \psi < 1$ is the share of early childhood education, and ϵ is the elasticity of substitution. Innate ability a and early childhood education h remain constant as an individual gets older. Individual productivity determines the effective labor supply of the individual. This labor contributes to the aggregate skilled or unskilled labor supply in the economy depending on the individual's job allocation.

The cost of endowing one's offspring with education level h in terms of the final good is given by the cost function

$$c_h(h, \Theta) = \kappa_h h^{\xi_h} \bar{z}^{\zeta/(\zeta+\lambda)} \tag{6}$$

where $\kappa_h > 0$ is a scale parameter, $\xi_h > 1$ introduces convexity, and $\bar{z}^{\zeta/(\zeta+\lambda)}$ ensures the cost scales up with aggregate output as the economy grows.

[7]Note that the $\bar{z}$ term in (3) introduces intertemporal spillover effects between the firms in the economy which is a salient feature of modern endogenous growth models. The additive structure is chosen over multiplicative because (i) it allows for solving the firm value functions in closed form and (ii) it ensures the existence of an invariant firm size distribution in a stationary equilibrium.

The innate ability of an individual is determined at the transition from childhood to young adult status, and depends on the innate ability of the parent. It is governed by a stochastic AR(1) process given by

$$\log a' = (1-\rho)\mu_a + \rho \log a + \epsilon_a, \ \epsilon_a \sim N(0, \sigma_a^2) \tag{7}$$

which has a mean of one.[8] The variables a and a' denote the innate ability of the parent and the child respectively. The persistence parameter ρ determines how much of the parental ability the child inherits. The stochastic innate ability shock ϵ_a is normally distributed with a mean of zero and variance of σ_a^2.

2.2.3 Tertiary education and job allocation

There are two types of jobs j in the economy: skilled/innovation jobs ($j = s$) and unskilled/production jobs ($j = u$). The job of an individual determines which pool his or her labor supply will contribute to, and hence the wage rate to be received per effective labor unit supplied (w_s if skilled and w_u otherwise). Any worker in the economy can get a production job. However, in order to get an innovation job, the individual needs to receive tertiary education at a high quality institution. This tertiary education provides the individual with (i) the training necessary to create innovations and (ii) an increase in the individual productivity, given by $l' = \Lambda l$.[9]

The ratio of high quality tertiary education available in the economy over total population is denoted by $\eta \in (0, 1)$ and assumed to be fixed.[10] Since the innovation jobs pay better than production jobs in equilibrium,[11] individuals would like to get innovation jobs. Because of this, high quality tertiary education is sought after; and since the supply is fixed, there is competition amongst individuals to receive it, which is cleared by a score mechanism described below.

At the beginning of the young adult period and after observing the innate ability a, each individual receives a score given by

$$\tilde{s}(l(h, a), n) = (1-\nu)l(h, a) + \nu n + \epsilon_j, \ \epsilon_j \sim N(0, \sigma_j^2) \tag{8}$$

[8] For a discussion on how well this parsimonious specification fits the intergenerational transmission of observed abilities, see Clark (2014).

[9] A good real world example of the described tertiary education would be an MSc or PhD degree in a STEM field at a high quality institution. Although a PhD is not always necessary to work on innovations, NSF National Survey of College Graduates (2003) shows that individuals who have invented at least one patent throughout their lifetime are three times more likely to hold a PhD degree compared to non-inventor college graduates.

[10] The opposite of this restriction would be having no restrictions on η, but fixing the score threshold $\bar{s}$ instead, so that any individual who has a sufficiently high score would get the high quality tertiary education. The quantitative experiments replicated with this alternative model deliver higher growth rate and welfare responses to parameter and policy changes; so the assumption to fix η puts discipline on the counterfactual implications of the model. See Appendix C.1 for details.

[11] This is not restriction of the model, but a result of the calibration exercise. See Section 4 for the details.

where $l(h,a)$ is individual productivity, n is credentialing spending (a choice variable), $\nu \in [0,1]$ is a parameter that governs the relative power of l versus n in determining the score, and ϵ_j is a normally distributed shock. After the scores for each individual are realized, the fraction η of the individuals with the highest scores receive high quality tertiary education, and are able to work in the innovation sector. The remaining $(1-\eta)$ fraction of the individuals do not receive high quality tertiary education and cannot create innovations, and thus have to work in the production sector.

In order to increase score upwards by the amount νn, the individual has to spend resources given by

$$c_n(n) = \kappa_n n^{\xi_n} \bar{z}^{\zeta/(\zeta+\lambda)} \tag{9}$$

in terms of the final good, where $\kappa_n > 0$ is a scale parameter, $\xi_n > 1$ introduces convexity, and $\bar{z}^{\zeta/(\zeta+\lambda)}$ ensures the costs scale up with aggregate output as the economy grows. This choice variable n captures any real world spending that increases the chances of getting a high quality tertiary education, such as hiring private tutors, preparing towards standardized tests, spending extra money to get into a college without a scholarship, the opportunity cost of studying as opposed to working in a job, etc.

Since the upper η fraction of the score distribution receives high quality tertiary education, there exists a score threshold $\bar{s}$ such that individuals with $\tilde{s} \geq \bar{s}$ receive the education, and the rest do not. In equilibrium, individuals with the necessary education always choose the innovation sector over the production sector, so the probability of getting high quality tertiary education and that of being a skilled worker are the same. The implied probability distribution of having job j for an individual is denoted by $F(j; l(h,a), n, \Theta)$. The aggregate state matters, since the score of a worker is only meaningful compared to the score threshold $\bar{s}$, as relative rank determines job allocation. The probability of having a skilled job is increasing in innate ability a, early childhood education h and credentialing spending n, whereas it is decreasing in the score threshold $\bar{s}$.

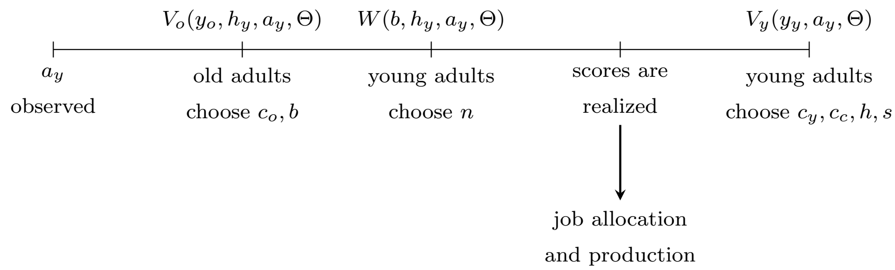
2.3 Maximization problems

2.3.1 Timing of events

Before moving on to the decision problems of the firms and the households, the timing of events within a period are listed below, which are also summarized on Figure 1:

1. The innate ability of young adults a_y is observed.
2. Old adults choose their bequests b and consumption c_o.
3. Young adults decide on credentialing spending n to receive a better score $\tilde{s}$.

FIGURE 1: TIMING OF EVENTS WITHIN A PERIOD.



4. Scores $\tilde{s}$ are observed, tertiary education is provided, and young adults are assigned to their jobs j.

5. Firms hire capital k and labor l_u and l_s for production and innovation. Production takes place and successful innovations are realized. Wages are paid.

6. Young adults choose how much to consume c_y, consumption of their children c_c, pre-tertiary education investment in their children h, and savings s.

2.3.2 Firm decision problems

The static profit maximization problem of the firm is given by

$$\Pi(z, \Theta) = \max_{k, l_u \geq 0} \{z^{\zeta} k^{\kappa} l_u^{\lambda} - (r + \delta)k - w_u l_u\} \tag{10}$$

where the firm pays interest rate plus depreciation $(r + \delta)$ and unskilled wage w_u for capital and unskilled labor services respectively. The associated capital and labor demand policy functions are denoted by $\hat{k}(z, \Theta)$ and $\hat{l_u}(z, \Theta)$.

Given the period profits $\Pi(z, \Theta)$ from the static maximization problem and the innovation technology described in (4), the intertemporal maximization problem of a firm can be written in recursive form as follows:

$$V(z, \Theta) = \max_{l_s \geq 0} \left\{ \Pi(z, \Theta) + \frac{\chi l_s^{\xi}}{1 + r} V(z + \gamma \bar{z}, \Theta') + \frac{(1 - \chi l_s^{\xi})}{1 + r} V(z, \Theta') - w_s l_s \right\} \tag{11}$$

The firm chooses how much skilled labor l_s to hire, which increases the likelihood of successful innovation χl_s^{ξ}. If successful, the firm's productivity next period is increased by $\gamma \bar{z}$. The prospect of earning higher profits in the future due to higher productivity provide incentives for the firm to engage in costly innovation. The skilled labor demand that solves this problem is denoted by $\hat{l}_s(z, \Theta)$.

2.3.3 Household decision problems

Given the ingredients of the model, there are three relevant decision problems for each household in any given period: (i) the bequest decision of old adults, (ii) the credentialing spending decision of young adults before job allocation, (iii) the consumption, pre-tertiary education investment and saving decisions of young adults after job allocation.[12] The associated value functions of the problems will be denoted by $V_o(.)$, $W(.)$ and $V_y(.)$ respectively.

2.3.4 Decision problem of the old

Let subscripts c, y and o stand for child, young and old respectively. Time subscripts will be suppressed for clarity. Let y denote wealth. Given the wealth of the old y_o, the early childhood education h_y and innate ability a_y of the young, and the aggregate state of the economy Θ, the bequest decision problem of the old can be stated as

$$\begin{aligned} V_o(y_o, h_y, a_y, \Theta) &= \max_{c_o, b \geq 0} \{u(c_0) + \alpha W(b, h_y, a_y, \Theta)\} \text{ s.t.} \\ c_o + b &\leq y_0 \end{aligned} \tag{12}$$

where c_o is the consumption of the old, b is the bequest left to the descendants and $\alpha > 0$ is the altruism parameter. Old agents choose how much bequests b to leave to their children who are now young adults, at the cost of reducing the their own consumption c_0. The problem is solved by the choice of a single variable b since preferences ensure the budget constraint holds with equality. Note the financial restriction that the bequests must be positive. This disallows agents from borrowing against the future income of their dynasty to consume today. The associated policy function is denoted by $\hat{b}(y_o, h_y, a_y, \Theta)$.

2.3.5 Decision problem of the young before job allocation

Given the bequest amount b, the early childhood education h_y and innate ability a_y of the young, and the aggregate state of the economy Θ, the credentialing spending decision problem of the young before job allocation can be stated as follows:

$$\begin{aligned} W(b, h_y, a_y, \Theta) &= \max_{n \geq 0} \{E\left[V_y(y_y, a_y, \Theta)|\cdot\right]\} \text{ s.t.} \\ y_y &= \left(w_{j_y} + \frac{w'_{j_y}}{1+r'}\right) l_y(h_y, a_y) + b - c_n(n) \\ j_y &\sim F(j; l_y(h_y, a_y), n, \Theta) \end{aligned} \tag{13}$$

[12]Children in a household have no decision problems to solve. They receive pre-tertiary education chosen by their parents and consume.

where j_y is a random variable that denotes job allocation and y_y stands for wealth as a young adult after job allocation. The wealth of the young y_y consists of the lifetime labor income and the bequests b received from parents, minus the cost of improving the score $c_n(n)$. The only choice variable is the resources spent on improving the score, denoted by n. Spending more resources increases the likelihood of getting a better job draw j_y distributed according to $F(j; l, n, \Theta)$ discussed earlier. The optimal n that solves this optimization problem is referred to as the credentialing spending policy function, $\hat{n}(b, h_y, a_y, \Theta)$.

Note that the young adults can borrow against their future lifetime labor income, so the model allows agents to borrow resources at the risk free interest rate r' to spend on credentialing which improves their chances of getting a high quality tertiary education. On the other hand, they cannot insure themselves against the idiosyncratic risk of not getting high quality tertiary education, which is always positive due to the shock term ϵ_j in (8). This forces them to be more prudent in increasing credentialing spending n by borrowing due to risk aversion.

2.3.6 Decision problem of the young after job allocation

Given the wealth y_y and the innate ability a_y of the young, and the aggregate state of the economy Θ, the consumption, early childhood education investment and saving decision problem of the young after job allocation can be stated as follows:

$$\begin{aligned} V_y(y_y, a_y, \Theta) &= \max_{c_y, c_c, h'_y, s \geq 0} \{u(c_y) + \alpha u(c_c) + \\ &\quad \beta E[V_o(y'_o, h'_y, a'_y, \Theta')|\cdot]\} \text{ s.t.} \\ y_y &\geq c_y + c_c + c_h(h'_y) + s \\ y'_o &= (1 + r')s \\ a'_y &\sim g(a_y) \\ \Theta' &= T(\Theta) \end{aligned} \tag{14}$$

Variables with primes indicate next period's values. The choice variables are the consumption of the young and their children, c_y and c_c, the early childhood education investment in the children h'_y which costs $c_h(h'_y)$ in terms of the final good, and the savings s. The sum of these expenditures must be below the wealth y_y. The expectation is over the innate ability a'_y of the child tomorrow, which depends on the innate ability of the parent a_y. The aggregate state of the economy evolves according to the transition function $T(.)$. The policy functions that solve this problem are given by $\hat{c}_y(y_y, a_y, \Theta), \hat{c}_c(y_y, a_y, \Theta), \hat{h}'_y(y_y, a_y, \Theta)$ and $\hat{s}(y_y, a_y, \Theta)$.

2.4 Balanced growth path equilibrium

Let $Z(z)$ denote the distribution of firm productivities in the economy. Labor market clearing implies

$$L_{u,t} \equiv \int_0^1 \hat{l}_{u,t}(z,\Theta)dZ(z) = 2(1-\eta)\int l(h,a)d\Phi_{u,t}(h,a), \text{ and} \tag{15}$$

$$L_{s,t} \equiv \int_0^1 \hat{l}_{s,t}(z,\Theta)dZ(z) = 2\eta\int l(h,a)d\Phi_{s,t}(h,a) \tag{16}$$

where $\Phi_{u,t}(h,a)$ and $\Phi_{s,t}(h,a)$ denote the joint distribution of early childhood education and innate ability at time t for unskilled and skilled workers respectively. The $(1-\eta)$ and η terms in the labor supply expressions are multiplied by average individual productivity because they designate the fraction of the population working in production and innovation sectors respectively. The terms are also multiplied by two since in any period both the young and old adults work. Aggregate savings in the economy is given by

$$A_{t+1} \equiv \int \tilde{a}_{m,t-1}d\tilde{A}(\tilde{a}) \tag{17}$$

where $\tilde{a}_{m,t} \equiv s_{m,t} - l(h_{m,t},a_{m,t})w_{j_{m,t},t+2}/(1+r_{t+2})$ denotes the net savings of the young adults of household m born at time t.[13] There are two kinds of assets in the economy: physical capital and shares in the bundle of firms $i \in [0,1]$. Both assets pay the risk-free interest rate r_t.[14] The capital market clearing requires the physical capital supply in the economy to equal the aggregate capital demand of the firms given by

$$K_t \equiv \int_0^1 \hat{k}_{u,t}(z,\Theta)dZ(z). \tag{18}$$

Final good market clearing requires

$$O_t = C_t + K_{t+1} - (1-\delta)K_t + N_t + H_t \tag{19}$$

where O_t denotes aggregate output and C_t, N_t and H_t are aggregate spending on consumption, score distortion and early childhood education investment at time t respectively. Finally, the number of people who receive high quality tertiary education must equal the exogenous restriction on their

[13] In order to calculate $\tilde{a}_{m,t}$, the labor income to be earned in the old adult stage is subtracted from $s_{m,t}$ because it was included in the expression y_y in the young agent's recursive decision problem. This was done to reduce the number of state variables to keep track of in the associated value function $V_y(.)$.

[14] Although each firm $i \in [0,1]$ faces idiosyncratic risk, aggregating over i makes profits received from the whole bundle a deterministic quantity due to the lack of aggregate fluctuations.

measure η. This imposes the condition

$$\eta = \int_{\bar{s}_t}^{\infty} \tilde{s} d\tilde{S}_t(\tilde{s}) \tag{20}$$

where $\tilde{S}_t(\tilde{s})$ is the score distribution at time t and $\bar{s}_t$ is the score cut-off above which agents get high quality tertiary education.

Given these ingredients, an equilibrium of this economy is defined as follows:

Definition 1 *An equilibrium is described by allocations* $[\{\vec{c}_{m,t}, b_{m,t}, n_{m,t}, h_{y,m,t}, s_{m,t}\}_{t=0}^{\infty}]_{m=0}^{1}$ *for households, allocations* $[\{z_{i,t}, k_{i,t}, l_{u,i,t}, l_{s,i,t}\}_{t=0}^{\infty}]_{i=0}^{1}$ *for firms, prices* $\{r_t, w_{u,t}, w_{s,t}\}_{t=0}^{\infty}$*, score cut-off* $\{\bar{s}_t\}_{t=0}^{\infty}$*, firm productivity distribution* $\{Z_t(z)\}_{t=0}^{\infty}$*, and joint distribution of jobs, early childhood education, and innate ability* $\{\Phi_t(j, h, a)\}_{t=0}^{\infty}$ *such that:*[15]

1. *Given prices and score cut-off, household allocations maximize* $V_o(b, h_y, a_y, \Theta)$, $V_y(y_y, a_y, \Theta)$*, and* $W(b, h_y, a_y, \Theta)$.
2. *Given prices and the productivity distribution, firm allocations maximize* $\Pi(z, \Theta)$ *and* $V(z, \Theta)$.
3. *All markets clear.*

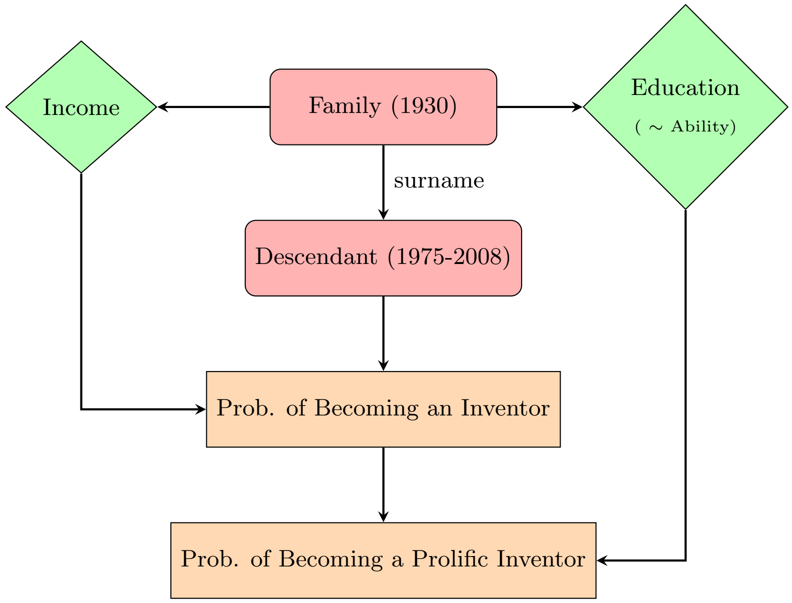
Output growth in this economy is driven by improvements in the productivities of the firms given by the distribution $Z_t(z)$. This paper focuses on the balanced growth path equilibrium where aggregate variables O_t, K_t, N_t, H_t, and C_t grow at the constant rate g. Along the balanced growth path, it turns out that the mean of the firm productivity distribution, $\bar{z} \equiv \int z dZ(z)$, is a sufficient statistic to determine the growth rate of the economy. Let the growth rate of the mean productivity $\bar{z}$ be denoted by g_z. Define transformed variables $\hat{z} \equiv z/\bar{z}^{\lambda/(\lambda+\zeta)}$, $\tilde{z} \equiv \bar{z}^{\zeta/(\lambda+\zeta)}$ and $\tilde{w}_s \equiv w_s/\tilde{z}$. The balanced growth path equilibrium of this economy is described below.

Theorem 1 *The balanced growth path equilibrium of the economy has the following form:*

1. *Aggregate allocations* O_t, K_t, N_t, H_t*, and* C_t*, and wages* $w_{u,t}$ *and* $w_{s,t}$ *grow at constant rate* g.
2. *Aggregate labor allocations* L_u *and* L_s*, interest rate* r*, score cut-off* $\bar{s}$*, and joint distribution of jobs, early childhood education, and innate ability* $\Phi(j, h, a)$ *are time-invariant.*
3. *Mean of the firm productivity distribution* $\bar{z}$ *grows at constant rate* g_z*, where* $1 + g = (1 + g_z)^{\zeta/(\lambda+\zeta)}$.
4. *Period profits of the firm are linear in* $\hat{z}$*, given by* $\Pi(z, \Theta) = \pi\hat{z}$.
5. *The value function of the firm is linear in* $\hat{z}$ *and* $\tilde{z}$*, given by* $V(z, \Theta) = v_1\hat{z} + v_2\tilde{z}$.

[15] Arguments of the allocations are suppressed for clarity.

Figure 2: Overview of the empirical analysis



6. *The constants v_1, v_2, π, prices $r, w_{u,t}, w_{s,t}$, growth rate g_z, and aggregate production factors K_t, L_u and L_s are jointly determined by a system of nonlinear equations given by (25), (26), (27), (29), (30), (31), and (32), and the market clearing conditions.*

Proof. See Appendix A ■

3 Empirical Analysis

3.1 Overview

In order to assess whether there is any indication of a misallocation of talent in innovation, several different data sources are combined. Figure 2 presents a simple schema of the baseline empirical analysis. The information on the probability of becoming an inventor, and how well one performs conditional on becoming one are obtained from various datasets that cover the years 1976-2008. The information on the family backgrounds come from the IPUMS-USA 5% sample of the U.S. census conducted in 1930. In order to link the recent patent and inventor micro-data to the older census data, surname information is used. Once the links between the families and the descendants are established at the surname level, the probability of becoming an inventor and the productivity as an inventor conditional on becoming one are regressed on family income and education. It is revealed that it is income and not education that predicts a positive probability of becoming an inventor, whereas it is education and not income that predicts the probability of becoming a prolific inventor. This inconsistency between the extensive and the intensive margins is the main focus of the empirical analysis. Following sections discuss the data sources in detail, describe the variables created, present and discuss the baseline empirical results, and conclude with some robustness checks.

3.2 Data Sources

3.2.1 NBER USPTO Utility Patents Grant Data

Patents are exclusionary rights, granted by national patent offices, to protect a patent holder for a certain amount of time, conditional on sharing the details of the invention. United States Patent and Trademarks Office (USPTO) is the agency in the U.S. Department of Commerce that issues patents to inventors and businesses for their inventions. From the great amount of information available in the files of USPTO, a substantial subsample has been compiled in an easy-to-use format by a group of researchers from the National Bureau of Economic Research (NBER) under the name NBER Patent Data Project (PDP).[16]

This dataset contains detailed information on 3,210,361 granted by the USPTO between the years 1976 and 2006. Each patent granted in the U.S. is assigned a unique patent number that makes it possible to link this dataset to many other datasets that contain information on patents some of which will be described further along. An important feature of this dataset is to provide citation links between individual patents. Similar to an academic paper, a new patent has to cite previous patents on which it builds, or other patents concerned with a similar but different invention, so that proper boundaries between the new and old patents can be established. The number of citations a patent receives from other patents is found in the literature to be a good proxy for its social and private value.[17] Since the citations a patent will receive throughout its lifetime cannot be known at a fixed point in time, and due to systematic citation differences between patents that belong to different technology classes, corrections need to be made to the citation numbers before using them as a proxy for patent quality. Hall, Jaffe, and Trajtenberg (2001) devise some correction weights to account for these biases, and their correction is used throughout this paper unless mentioned otherwise.

3.2.2 The Careers and Co-Authorship Networks of U.S. Patent-Holders

Filing a patent application in the U.S. requires providing the names of three types of individuals in the application form: The assignees who own the patent once granted; the applicants who are responsible for legal correspondence with USPTO; and the inventors who actually came up with the

[16]For more information, please visit `https://sites.google.com/site/patentdataproject/`

[17]For instance, Hall, Jaffe, and Trajtenberg (2005) argue that the citation-weighted patent portfolio of a firm is a plausible indicator of the intangible knowledge stock of a private firm, and that this measure has additional explanatory power for the market value of the firm beyond the conventional discounted sum of R&D spending.

innovation.[18] Extensive information on the inventors of patents granted in the U.S. between years 1975-2008 is obtained from a dataset produced by Lai, D'Amour, and Fleming (2009).[19] Unlike the PDP data, this dataset contains the names of every inventor who has worked on a patent granted in the U.S. between years 1975-2008. This is crucial, since 55.3% of the patents in the sample were created by a group of inventors. Another novel feature of this data is the provision of a unique inventor identifier which makes it possible to track the patent portfolio of individual inventors throughout their careers.

The dataset contains 8,031,908 observations at the patent × inventor level, and 2,229,219 unique inventors. Among other variables, the dataset contains address information of the inventors as well as their names and surnames. The address information is used to determine the country the inventor lives in at the registration date of the patent so that the analysis can be restricted to U.S. inventors only. The surname information is used to create a relative representation (among inventors) measure at the surname level and link the socioeconomic background data from 1930 to inventors today. Both of these will be discussed in detail.

3.2.3 IPUMS-USA 1930 5% Sample

Integrated Public Use Microdata Series (IPUMS-USA) is a project dedicated to collecting and distributing United States census data, and it consists of more than fifty high-precision samples of the American population drawn from federal censuses. The particular sample used in this project is the 1930 sample which contains information on 5% of all Americans who were counted in the 1930 census. The 1930 sample is preferred over other samples since it is the most recent publicly available sample that contains name and surname information at the observation level.[20]

Since the dataset contains census information, the wealth of information at the individual level is immense. The main information derived from this dataset is on socioeconomic status of people with a particular surname, such as income and education collapsed at the surname level. Similar to other studies that use the IPUMS samples prior to 1940 (Olivetti and Paserman, 2013), income associated with a surname is measured using the OCCSCORE variable measured in hundreds of 1950 U.S. dollars. This variable includes income from non-wage activities such as interest income and dividends in addition to earnings. Finally, EDSCOR50 variable is used as the education variable

[18] Hence, the owner of a patent or the manager in an innovating firm are not listed as inventors unless they took part in the innovation process. USPTO explicitly states the following: "All inventor(s) named in the provisional application must have made a contribution, either jointly or individually, to the invention disclosed in the application."

[19] Please visit `http://hdl.handle.net/1902.1/12367` to access the data.

[20] Individual questionnaires of any specific census are not released by the National Archives until 72 years after that specific census has been taken due to confidentiality requirements. Name and surname information is also available for other samples spanning the years 1850-1920 in the IPUMS database; however they are less recent, and most of these samples are at 1% level instead of 5%.

which measures college attendance. Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek (2010) and the project website contain further details on the dataset and variables.[21]

3.2.4 Demographic Aspects of Surnames from Census 2000

This dataset released by the U.S. Census Bureau in 2007 contains information on the overall frequency of surnames in the U.S. constructed using the 2000 decennial census of population, based on approximately 270 million individuals with valid surnames.[22] It contains 151,671 unique surnames. Combined with the U.S. inventor data previously discussed, it is possible to create measures of probability of becoming an inventor at the surname level. This dataset further includes information on the ethnicity distribution for each surname broken down into six categories (White, Black, Hispanic, Asian or Pacific Islander, American Indian or Alaskan Native, or mixed). These variables are used to create dominant race fixed effects for race associated with a surname. One caveat of the data is that it only includes surnames that have a frequency above hundred, which makes it unsuitable to use in questions regarding extremely rare surnames. Such rare surnames are therefore excluded from the following analysis.

3.3 Data construction and variables

The summary statistics for the variables used in the empirical analysis are presented on Table 1. The following subsections describe what they stand for and how they are generated.

3.3.1 Surname level socioeconomic status variables (1930)

Socioeconomic status variables such as income, earnings, and education are constructed at the surname level by taking the averages of observations in the IPUMS-USA 1930 5% sample. In this process, observations without a valid occupation are not included.[23]

3.3.2 Relative representation of a surname among inventors (1975-2008)

The extensive margin analysis focuses on the question of how the socioeconomic background of an individual affects the probability of becoming an inventor – or using the model's terminology, the probability of being assigned to a job in the innovation sector. The Careers and Co-Authorship Networks of U.S. Patent-Holders data contains the names of all inventors who worked on patents granted in the U.S. between the years 1975 and 2008, from which it is possible to obtain the number of inventors with a particular surname. However, the fact that there are many inventors with the

[21] IPUMS-USA project website can be accessed at `https://usa.ipums.org/usa/index.shtml`.

[22] Refer to Word, Coleman, Nunziata, and Kominski (2008) for a detailed description, and `http://www.census.gov/genealogy/www/data/2000surnames/index.html` for the data.

[23] These observations correspond to those with OCC1950 values between 980 and 999. Visit `https://usa.ipums.org/usa-action/variables/OCC1950#codes_section` for a complete list of OCC1950 values.

Table 1: Summary Statistics

	Observation	*Mean*	*St. Dev*
Panel A. Extensive Margin Analysis			
relative representation (1975-2008)	110,290	83.1	68.6
income (1930)	110,290	21.7	3.70
education (1930)	110,290	15.8	17.5
is black	110,290	2.28	14.9
is asian	110,290	1.52	12.2
is native	110,290	.048	2.20
is hispanic	110,290	11.4	31.8
is mixed	110,290	.002	.438
Panel B. Intensive Margin Analysis			
log quality weighted total patents (1975-2006)	81,348	3.65	.664
log average patent quality (1975-2006)	81,348	2.53	.387
log maximum patent quality (1975-2006)	81,348	2.95	.455
log total patents (renewed thrice) (1975-2006)	78,438	.695	.287
log total patents (top 10% only) (1975-2006)	81,348	.321	.200
log income (1930)	81,348	3.11	.158
log education (1930)	81,348	2.69	.494
is black	81,348	2.05	14.2
is asian	81,348	1.55	12.3
is native	81,348	.035	1.87
is hispanic	81,348	11.7	32.1
is mixed	81,348	.001	.38

Notes: Relative representation and dominant race indicator variables are multiplied by 100 for clarity. The means and standard deviations reported on the table are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population of 2000. Patent quality is measured by the number of patent citations corrected for truncation using the correction terms from Hall, Jaffe, and Trajtenberg (2001).

surname Smith does not mean that Smiths are more likely to become inventors by itself. In order to create a measure of the probability, the number of inventors with a particular surname is divided by the number of all people in the U.S. with the same surname obtained from Demographic Aspects of Surnames from Census 2000; i.e.

$$\text{inventor probability (surname)} = \frac{\text{number of inventors (surname)}}{\text{number of individuals (surname)}}$$

Relative representation of a surname among the inventor sample is then built simply by dividing the inventor probability associated with the surname with the unconditional probability of becoming an inventor in the U.S. given by

$$\text{relative representation (surname)} = \frac{\text{inventor probability (surname)}}{\text{unconditional inventor probability}}$$

Thus a relative representation score above unity means that individuals with that surname are more likely to become inventors than the average person, and vice versa.

3.3.3 Patent and inventor quality metrics (1975-2006)

The intensive margin analysis considers the question of how the socioeconomic background of an individual affects the productivity as an inventor conditional on becoming one. In order to conduct this analysis, it is necessary to come up with metrics that measure inventor productivity. The unique inventor variable allows tracking the patent portfolio of each inventor between the years 1975 and 2008. The productivity of an inventor can be calculated as a function of the information on all the patents he or she has worked on. This naturally leads to the question of how to assess the value of a patent. In line with the literature, the quality of a patent is proxied by the citations received by the patent, corrected for truncation bias and other concerns using the weights devised by Hall, Jaffe, and Trajtenberg (2001). The patent quality information from the PDP data is linked to the inventor data using the unique patent numbers granted by USPTO. The inventor quality metric that is used in the baseline analysis is the total quality weighted patents of an inventor throughout his or her career. Several additional alternative metrics are considered in the robustness analysis in Section 3.4.3.

Since the data contains all inventors who worked on patents registered in the U.S., it is necessary to separate the foreign inventors from the sample used to create surname level variables. The address information of an inventor is available for every patent, and there is considerable variation between the countries. For this study, only inventors who have stayed in the U.S. throughout their whole career are kept. Average inventor quality metrics at the surname level are constructed by taking the unweighted average of individual inventor qualities.

3.4 Empirical results

3.4.1 Extensive margin analysis

In order to understand whether there is a misallocation of talent in the innovation sector or not, it is necessary to empirically demonstrate what is correlated with the probability of having a job in this sector. The surname level probability of being an inventor is used as a proxy to gauge this, although inventors are not the only individuals who work in the innovation sector. Socioeconomic background information at the surname level obtained from IPUMS-USA 1930 dataset is connected to these probabilities using surnames. Standard ordinary least squares estimation is used where the

Table 2: Extensive Margin Baseline

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.239***		.239***
	(.010)		(.010)
education (1930)		.029***	.001
		(.006)	(.005)
Obs.	110,290	110,290	110,290
R^2	0.27	0.23	0.27

Notes: Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

relative representation rate is regressed on the socioeconomic variables: income and education.[24] The three columns of Table 2 correspond to regressions on income, education, and both variables at the same time respectively.

Looking at the first two columns, it is observed that both income and education associated with a surname in 1930 are positively correlated with the relative representation among inventors today (1975-2008) and statistically significant. A standard deviation increase in income increases the relative representation rate by 23.9% compared to its standard deviation, while a standard deviation increase in education increases it by 2.90%. Given that there are roughly three generations between 1930 and today, these numbers are quite substantial ($\sqrt[3]{23.9\%} = 62.1\%$), and hint towards low intergenerational mobility in social status, similar to the results in other studies that use surnames (Clark, 2014).

Looking at the third column tells another striking story: It is income and not education that is strongly correlated with the over-representation among inventors. In other words, people with surnames that were richer in the past are more likely to become inventors today; but controlling for income, education has no further prediction power.[25] This finding is the motivation behind the inclusion of the credentialing spending in the model, which enables agents to increase the

[24] Since all of the variables are averages calculated at the surname level, using an ordinary least squares estimator is essentially identical to a two-sample two-stage least squares estimator. Alternatively, if an observation was created for each individual in the Census 2000 sample with an indicator variable for being an inventor or not, the OLS regression of this indicator variable on background information linked by surnames would yield the same results as Table 2.

[25] The insignificance of education is not driven by multicollinearity due to high correlation between the variables. In order to address such concerns, a variance inflation factor test is conducted after each regression in the paper in which both income and education are included as regressors. None of the tests result in a VIF large enough to be concerned about (< 3). The results are available upon request.

probability of getting the high quality tertiary education necessary for innovation sector jobs by spending private resources.

3.4.2 Intensive margin analysis

Table 3: Intensive Margin Baseline

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	.066***		.001
	(.009)		(.009)
log education (1930)		.176***	.175***
		(.008)	(.009)
Obs.	81,348	81,348	81,348
R^2	0.03	0.05	0.05

Notes: Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Having discovered that income associated with the surname is significantly positively correlated with the probability of being an inventor, the natural next step is to ask whether these individuals are the individuals who would make the best inventors. In order to investigate this question, the inventor quality metric described earlier is regressed on income and education. Table 3 displays the results of three OLS regressions: log inventor quality on log income, on log education, and on both variables at the same time.

Due to the log-log specification, the coefficients can be interpreted as elasticities. By themselves, both income and education turn out to be positively correlated with inventor quality, and the associated coefficients are statistically significant. Once again, given that there are three generations between the samples, the elasticity estimates are considerably high. However, this time the standard error for education is much smaller than that for income, the opposite of what was observed in the extensive margin analysis.

The last column regresses inventor quality on both income and education, and the results are striking. The elasticity of inventor quality with respect to education is very close to that on column 2, but the elasticity with respect to income vanishes, and is statistically insignificant. Conditional on becoming an inventor, it is the inventors with “more educated” surnames who are the most successful in creating new path-breaking innovations. This is in direct contrast to the extensive margin results, and suggests that the individuals who would make the best inventors might not be the same as those the society allocates as inventors. This fact is captured in the model by three ingredients: (i) education increases individual productivity in the innovation sector, (ii) education and innate ability are complementary in determining individual productivity, (iii) credentialing spending increases the probability of getting in an innovation sector job, but it does not increase individual productivity compared to other inventors (as opposed to education, which does both).

3.4.3 Alternative inventor quality measures

Table 4: Intensive Margin Robustness - Alternative Measures

	log avg. patent quality (1975-2006)	log max. patent quality (1975-2006)	log total patents (renewed thrice) (1975-2006)	log total patents (top 10% only) (1975-2006)
log income (1930)	.000	.013*	.031***	.033***
	(.008)	(.008)	(.009)	(.008)
log education (1930)	.130***	.142***	.098***	.085***
	(.008)	(.008)	(.008)	(.008)
Obs.	81,348	81,348	78,438	81,348
R^2	0.02	0.04	0.05	0.03

Notes: See notes for Table 3.

In the baseline intensive margin analysis, quality weighted total patents of an inventor was used as the inventor quality metric, where patent quality was measured by the citations a patent receives. This section establishes that the results are robust to using different measures of inventor quality. Results pertaining to additional alternative measures can be found on Table 12 in the empirical appendix.

Table 4 replicates the regression on column 3 of Table 3 using different inventor quality metrics.[26] The first two columns preserve the same patent quality metric (citations), but consider the average and maximum patent quality for inventors respectively. Compared to the baseline measure, the average patent quality measure puts less weight on inventors who come up with a high number of innovations which are of mediocre quality. Similarly, the maximum patent quality measure only considers the best invention of a given inventor, comparing inventors according to the best ideas they came up with and ignoring everything else. The results are very similar to the baseline analysis: log education dominates in both regressions, and log income is either statistically insignificant, or significant at the 10% level and economically insignificant.

[26]Results for replicating columns 1 and 2 are also very similar, but excluded to conserve space. These are available upon request from the author.

On column 3, a new patent quality metric is introduced: patent renewal status. USPTO requires the patent holders to renew their patents on the 4th, 8th and 12th years after the patent grant date by paying a small fee. If the patent holders do not renew their patents on these dates, they lose the monopoly rights on their invention. There is significant variation about how many times patents are renewed. The patent quality metric used on column 3 assigns a quality of 1 if the patent was renewed three times throughout its duration, and 0 otherwise. Hence only patents which were seen sufficiently valuable by their holders to renew three times are counted.[27] The results with this metric are similar in that education dominates income, but this time the effect of income is not statistically insignificant.

Last column does the opposite, and focuses on a patent quality measure that only puts weight on the best inventions produced in a year. For each year, the patents are ranked according to the citations they receive. Only the top 10% of the inventions in a given year are assigned a quality of 1, whereas the remaining 90% are assigned a quality of 0. Using the inventor quality measure derived from this new measure of patent quality, the results are similar to column 3: education is found to dominate income once again.

3.4.4 Controlling for demographic changes and immigration between samples

The Unites States is a country of immigrants, and it has received significant immigration during the time period from 1930 to 2008. Many surnames that were very rare in the 1930s are now quite common. In contrast, some surnames are now less frequent, either due to being crowded out by the new or existing surnames, or due to low number of offspring or higher mortality rates. Could any of these demographic changes bias the obtained estimates in a particular direction, potentially causing wrong conclusions to be drawn? Recognizing this possible problem, this section is dedicated to investigating whether this is true.

In order to tackle this issue, a simple variable called population share ratio is constructed. The share of a surname in the population in 2000 is divided by that in 1930. This ratio is larger than unity if the surname has increased in frequency, which is the case for many immigrant surnames. Conversely it is smaller than unity for surnames which actually lost their prominence over time. Using this ratio as an additional explanatory variable, Table 5 repeats the extensive margin regression on column 3 of Table 2. Columns 1 and 2 repeat the regression after dropping the top and bottom 25% of the sample according to population share ratio. Hence they drop the extremely over- and under-achieving surnames from the sample respectively. Columns 3 and 4 repeat the same exercise keeping only the top and bottom halves of the sample respectively, i.e. looking at over- and

[27] Note that although this quality metric is very reliable in weeding out patents that turn out to be worthless over time, it provides no quality variation between patents which are sufficiently valuable to be renewed every single time. Hence it should be thought of as a quality measure that is more informative in the lower tail of the unobserved quality distribution as opposed to the upper tail.

under-achievers within their own groups. The last column retains the whole sample, but includes the population share ratio as a linear regressor. Although the magnitudes change, income is found to be dominant in all cases, whereas education is found to be either insignificant, or significant but negatively correlated. In addition, when included as a linear regressor, the population share ratio turns out to be insignificant. Consequently, the findings of the extensive margin analysis are found to be robust.

TABLE 5: IMMIGRATION ROBUSTNESS (1930-2000) - EXTENSIVE MARGIN

	relative representation (1975-2008)				
	(1)	(2)	(3)	(4)	(5)
income (1930)	.341***	.249***	.232***	.158***	.239***
	(.009)	(.011)	(.013)	(.006)	(.010)
education (1930)	−0.021***	.003	.005	−.002	.001
	(.004)	(.005)	(.006)	(.004)	(.005)
pop. share(2000)/pop. share(1930)					−.025
					(.032)
Obs.	82,718	82,735	55,148	55,210	110,290
R^2	0.13	0.30	0.34	0.04	0.27

NOTES: COLUMNS 1 AND 2 REPEAT THE REGRESSION IN THE LAST COLUMN OF TABLE 2 AFTER DROPPING THE TOP AND BOTTOM 25% OF THE SAMPLE ACCORDING TO POPULATION SHARE RATIO RESPECTIVELY. COLUMNS 3 AND 4 REPEAT THE SAME EXERCISE FOR THE TOP AND BOTTOM HALVES OF THE SAMPLE RESPECTIVELY. COLUMN 5 REPEATS THE SAME REGRESSION WITH THE WHOLE SAMPLE WHILE INTRODUCING THE POPULATION SHARE RATIO LINEARLY AS A REGRESSOR IN ADDITION TO INCOME AND EDUCATION. ALL NOTES FOR TABLE 2 APPLY.

Table 6 repeats the same analysis done in Table 5 for column 3 of Table 3. The results are quite similar: Although the exact quantitative magnitudes may vary, the effect of education is always quite large and positive, dominating that of income. The effect of income is found to be either statistically or economically insignificant in all cases. The only difference is observed when population share ratio is added as a linear regressor: Its coefficient turns out to be significant at the 5% level and positive. However, estimated at 1.6%, its coefficient is much smaller compared to the coefficient of education (17.7%). Hence, it is once again concluded that the findings in the intensive margin analysis are robust.

One could also be worried about another issue: It is possible that a surname the frequency of which is stable over the 1930-2008 time period actually belonged to people who were recent immigrants in 1930. Systematic differences between such surnames and those who were already largely stable in frequency prior to 1930 could lead to potential biases similar to those discussed earlier. Luckily, it is possible to construct a similar population share ratio using surname frequencies in 1930 and 1880, using an earlier IPUMS-USA sample. The cost of doing so is losing observations

Table 6: Immigration Robustness (1930-2000) - Intensive Margin

	log quality wtd. total patents (1975-2006)				
	(1)	(2)	(3)	(4)	(5)
log income (1930)	$-.015^{*}$	.011	.011	$.037^{***}$	.001
	(.009)	(.010)	(.012)	(.008)	(.009)
log education (1930)	$.162^{***}$	$.173^{***}$	$.177^{***}$	$.145^{***}$	$.177^{***}$
	(.006)	(.010)	(.012)	(.008)	(.009)
pop. share(2000)/pop. share(1930)					$.016^{**}$
					(.032)
Obs.	61,013	61,011	40,684	40,676	81,348
R^2	0.03	0.06	0.07	0.03	0.05

Notes: Columns 1 and 2 repeat the regression in the last column of Table 3 after dropping the top and bottom 25% of the sample according to population share ratio respectively. Columns 3 and 4 repeat the same exercise for the top and bottom halves of the sample respectively. Column 5 repeats the same regression with the whole sample while introducing the population share ratio linearly as a regressor in addition to income and education. All notes for Table 3 apply.

that belong to surnames which do not exist in the 1880 census sample. The results of this robustness analysis are qualitatively very similar, and can be found on Tables 13 and 14 in the empirical appendix.

3.4.5 Summary of empirical results

The two stylized facts obtained in the empirical analysis can be summarized as follows:

Fact 1: Individuals from richer backgrounds are much more likely to become inventors (23.9%); whereas those from more educated backgrounds experience no similar advantage (0.1%).

Fact 2: Conditional on becoming an inventor, individuals from more educated backgrounds turn out to be much more prolific inventors (17.5%); whereas those from richer backgrounds exhibit no such aptitude (0.1%).

However, these results by themselves would be insufficient to establish whether there is an economically significant misallocation of talent or not, given that innate ability is unobserved in the data. This is important, since (i) innate ability is likely to play a large role in determining the probability of becoming an inventor as well as success conditional on becoming one, (ii) innate ability is found to be very persistent across generations by other studies (Clark, 2014), and this might be causing the observed strong positive correlations. In order to measure the extent of the misallocation of talent in innovation, the model developed in Section 2 is employed, where the regressions run here are replicated within the model, targeting the empirical coefficient estimates. The next section describes this calibration exercise.

4 Calibration

4.1 Solution method

Computation of the solution requires value function iteration to solve for $V_o(y_o, h, a; \Theta)$, $W(b, h, a; \Theta)$ and $V_y(y_y, a; \Theta)$ and the associated policy functions $\hat{b}(y_o, h, a; \Theta)$, $\hat{n}(b, h, a; \Theta)$, $\hat{h}(y_y, a; \Theta)$ and $\hat{s}(y_y, a; \Theta)$. Simulation of the joint stationary distribution of jobs, innate ability, and early childhood education as well as the stationary distribution of normalized savings are necessary to calculate the aggregate supplies as well as the cut-off score threshold $\bar{s}$. The results of the firm's maximization problem and the market clearing conditions boil down to analytical non-linear equations in K, L_u and L_s as discussed in Section 2. Then these are solved to obtain the balanced growth path equilibrium. The pseudo-code for the algorithm used to solve for the BGP equilibrium can be found in Appendix A.

4.2 Identification

The simulation of the model requires the assignment of values to several parameters. There are nineteen parameters to pick: $\beta, \omega, \alpha, \kappa, \lambda, \delta, \Gamma, \xi, \psi, \epsilon, \rho, \nu, \sigma_a, \eta, \kappa_h, \xi_h, \kappa_n, \xi_n, \sigma_j$. In order to select values for the parameters, a set of empirical targets are specified for the model to match. Some common parameters are chosen from existing studies, and the rest are internally calibrated by employing a minimization routine that seeks to match the data targets with the associated model-generated counterparts. In particular, some of the regressions found on Section 3 are replicated in the model, and the minimization algorithm attempts to achieve the same coefficients ("betas") with regressions run on model-simulated data, where the variables are normalized in the same manner. The summary of the calibration exercise is presented on Table 7. The details are as follows:

1. *CRRA parameter:* This parameter is taken to be $\omega = 2.00$, consistent with the estimates listed in Kaplow (2005).

2. *Parental altruism parameter:* This variable is chosen to be $\alpha = 0.50$, following Aiyagari, Greenwood, and Seshadri (2002).

3. *Capital's and labor's share of income:* Corrado, Hulten, and Sichel (2009) calculate the shares of tangible capital, labor, and intangible capital to be $\kappa = 0.25$, $\lambda = 0.60$ and $\zeta = 0.15$ respectively. The share of intangible capital they calculate is mapped to the share of productivity of a firm in generating output in the model.

4. *Depreciation rate for capital:* The annual depreciation rate of physical capital is chosen as 6.9% which is consistent with the U.S. National Income and Product Accounts. Since each period lasts 25 years, $\delta = 0.82$.

5. *Concavity of innovation production:* Following Hall and Ziedonis (2001), the concavity parameter of the innovation production function is chosen as $\xi = 0.50$. This is the most widely

Table 7: Parameter Values

Parameter	*Description*	*Identification*
	External Calibration	
$\omega = 2.00$	CRRA parameter	Kaplow (2005)
$\alpha = 0.50$	Parental altruism	Aiyagari, Greenwood, and Seshadri (2002)
$\kappa = 0.25$	Capital's share in production	Corrado, Hulten, and Sichel (2009)
$\lambda = 0.60$	Labor's share in production	Corrado, Hulten, and Sichel (2009)
$\delta = 0.82$	Depreciation rate	U.S. NIPA
$\xi = 0.50$	Concavity of innovation production	Hall and Ziedonis (2001)
$\sigma_a = 0.70$	St. dev. of innate ability shock	Knowles (1999)
$\eta = 11.6\%$	Fraction of skilled jobs	U.S. Census (2013)
	Internal Calibration	
$\beta = 0.28$	Discount factor	Real interest rate
$\Gamma = 0.92$	Innovation productivity increase	GDP growth rate
$\rho = 0.70$	Persistence of innate ability	IG corr. of earnings
$\kappa_h = 0.04$	Cost of pre-tertiary education investment	Education spending/GDP
$\kappa_n = 0.05$	Cost of score distortion investment	Inequality targets
$\xi_h = 1.30$	Convexity of pre-tertiary education inv.	Inequality targets
$\xi_n = 2.50$	Convexity of score distortion inv.	Inequality targets
$\psi = 0.40$	Education share of ind. productivity	Regression targets
$\epsilon = 1.90$	Ind. productivity elasticity	Regression targets
$\nu = 0.89$	Influence of credentialing spending	Regression targets
$\sigma_j = 0.80$	St. dev. of job shock	Regression targets

Notes: All internally calibrated parameters are identified jointly; the moments in the internal calibration panel are provided for intuition.

used value in the literature.

6. *Standard deviation of innate ability shock:* This parameter is chosen to be $\sigma_a = 0.70$, in line with findings on empirical income distributions reported in Knowles (1999).

7. *Fraction of skilled jobs:* This parameter is chosen such that it equals the percentage of individuals in the U.S. with graduate degrees, which is 11.6% (U.S. Census, 2013).

8. *Long-run interest rate:* The long-run interest rate of 4.0% is targeted, which determines the discount factor β.

9. *Long-run output growth:* Since 1945, the the aggregate output in the U.S. grew at circa 2% per year. The parameter Γ determines the increase in productivity innovation generates, and hence it plays the foremost role in determining the output growth rate in the model.

10. *The ratio of education spending to GDP:* The ratio of the aggregate spending on education to GDP in the U.S. is around 7.30% (OECD, 2013).The model counterpart of this ratio is the aggregate resources spent on education over total output.

Table 8: Calibration Targets

Target	*U.S. Data*	*Model*
Aggregate targets		
Yearly real interest rate	4.00%	4.00%
Yearly GDP growth rate	2.00%	2.00%
Education spending/GDP	7.30%	8.55%
Intergenerational correlation targets		
IG corr. of earnings	0.70	0.70
IG corr. of wealth	0.37	0.33
Inequality targets		
Wage income Gini index	0.48	0.52
Log 90/10 ratio	1.08	1.17
Log 90/50 ratio	0.46	0.52
Log 50/10 ratio	0.62	0.65
Regression targets		
Extensive margin, income effect	0.24	0.19
Extensive margin, education effect	0.00	0.07
Intensive margin, income effect	0.00	0.08
Intensive margin, education effect	0.18	0.22

11. *Intergenerational correlation of earnings:* The persistence of earnings across generations is an important statistic for the model to replicate, since it puts discipline on the persistence of innate ability which is unobserved. The value of 70% is targeted in the baseline analysis (Knowles (1999)).[28]

12. *Intergenerational correlation of wealth:* The persistence of wealth across generations is also an important statistic to replicate, since the mechanism that generates the misallocation of talent in the model works through the wealth inequality between households. This value is estimated to be 37% in Charles and Hurst (2003).

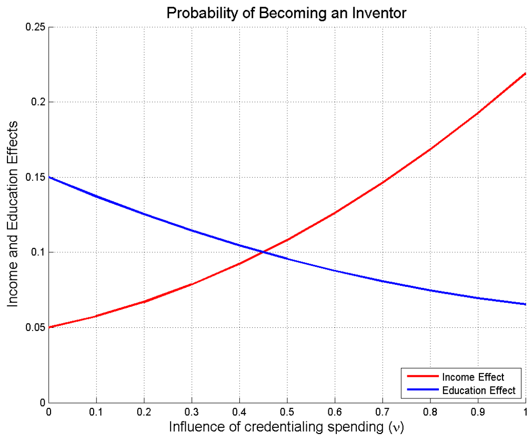
13. *Inequality targets:* The calibration procedure aims to generate a realistic income distribution. To this end, various inequality metrics are calculated using the model-generated distribution, and matched with their empirical counterparts. These are the Gini index, and log 90/10, 90/50 and 50/10 ratios.[29]

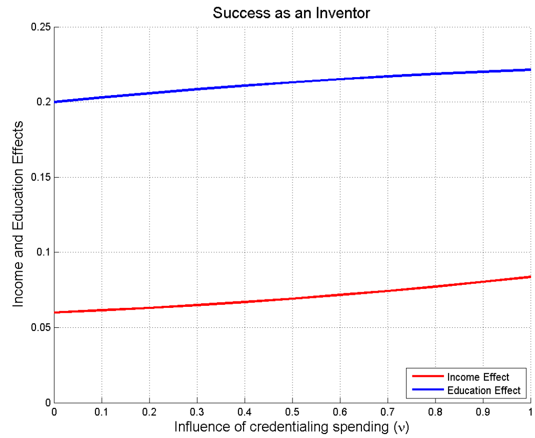
14. *Indirect inference:* The baseline extensive and intensive margin regressions in Section 3 are replicated in the model. Income is proxied by the income of the agents in the model, and education is proxied by pre-tertiary education. Relative representation among inventors in

[28]Since there are also estimates of intergenerational correlation of earnings as low as 40% in the literature, the model is re-estimated with a lower target as a robustness check in Section 5.4.

[29]Note that one of these three ratios is a deterministic function of the other two; so it provides no additional information.

FIGURE 3: CHANGES IN INCOME AND EDUCATION EFFECTS WITH VARYING VALUES OF ν





the data is mapped to relative representation in the innovation sector jobs. Inventor quality in the data is mapped to individual productivity conditional on having an innovation job. As in the empirical analysis, all variables are normalized by subtracting the mean and dividing by the standard error. The coefficients of income end education in both margins are then targeted.

The success of the calibration exercise in matching the data targets is presented on Table 8. The interest rate and the yearly GDP growth rate are hit very precisely, and they determine the values of β and Γ respectively. The model generates an education spending to GDP ratio somewhat higher than what is observed in the U.S. data. Given that the number taken from the data does not include the opportunity cost of time spent by parents in order to nurture their children, overshooting might not be a significant problem.

The intergenerational correlation of earnings is hit precisely, which disciplines the persistence of (unobserved) innate ability ρ (positively related), but is also influenced by the standard deviation of the idiosyncratic job shock σ_j (negatively related). The intergenerational correlation of wealth the model produces is 0.33, which is somewhat lower than the value of 0.37 observed in the data, but still within a reasonable range.

The model generates a wage income distribution slightly more unequal compared to the U.S. economy. For instance, the Gini index is calculated to be 0.52 as opposed to 0.48 observed in the data. However, the remaining inequality targets that measure the inequality in different sections of the distribution show that the model is successful in matching the shape. Log 90/10, log 90/50 and log 50/10 ratios are all slightly higher than their data counterparts by similar percentages.

The model is able to replicate the dominance of income on the extensive margin (the probability of getting an innovation sector job) and the dominance of education on the intensive margin (the success conditional on getting an innovation job). The coefficients of the dominated effects (education on the extensive margin, and income on the intensive margin) are not precisely zero, so the starkness of the differences are more similar to those observed in the regressions on Columns 3 and 4 of Table 4, as opposed to that on Column 3 of Table 3 on the intensive margin.

Generating the discrepancy between the effects of income and education on the two margins is made possible by the credentialing spending channel. Figure 3 plots the effects of income (red) and education (blue) while varying the influence of credentialing spending ν in the range of values it can take ($\nu \in [0, 1]$). The left panel plots the effects on the extensive margin, and the right panel plots the same on the intensive margin. As ν increases from 0 to 1, the predictive power of ancestor income on the probability of becoming an inventor increases, whereas that of education decreases. On the other hand, increasing ν from 0 to 1 does not change the predictive power of ancestor income and education in opposite directions, slightly increasing both at the same time.[30] This makes it possible to change the value of ν such that the dominance pattern observed in the data can be hit in the model generated regressions. This differential effect of ν on the two margins provides the intuition on how targeting the dominance pattern helps pin down its value.[31]

5 Quantitative Results

In this section, using the parameter values estimated in Section 4, several quantitative experiments are conducted to better understand the mechanism of the model, to assess the welfare costs associated with the misallocation of talent due to the credentialing spending channel, and to determine socially optimal progressive bequest tax schedules.

The first subsection describes the social welfare function used in the study, and how two different steady-state economies are compared with each other. The following subsection conducts a hypothetical thought experiment where the credentialing spending channel is completely shut down, which results in an increase in the aggregate output growth rate as well as social welfare through a reduction in the misallocation of talent.

The third subsection focuses on how a benevolent government can increase social welfare and economic growth in a decentralized market economy through the policy tool of progressive bequest

[30] Whether income or education dominates on the intensive margin (i.e. inventor productivity) is determined by other parameters of the model. The model is able to generate any correlation pattern, including the exact opposite of the empirically observed pattern of dominance, by changing the parameter values.

[31] One could be worried about whether other parameters that play a part in the determination of the individual score could generate a similar differential effect on the two margins. The prime candidates are the elasticity of substitution between innate ability and early childhood education ϵ, and the share of education in individual productivity ψ. It is found out that this is not the case. In particular, if the credentialing spending channel is shut down ($\nu = 0$), both ϵ and ψ change the effects of income and education in the same direction on both margins at the same time. Therefore if the credentialing spending channel is removed from the model without introducing any other mechanisms, the model is unable to mimic the domination patterns observed in the data.

taxation. Although the growth effect is found to be around 25% of what can be achieved by shutting down the credentialing spending channel, the welfare increase is found to be larger and quite significant: 6.20% in consumption equivalent terms

In Section 5.4, the model is recalibrated with a lower intergenerational correlation of earnings target of 0.45 in order to check whether the model generates similar quantitative implications. Repeating the credentialing spending shut-down experiment with the new calibration leads to higher growth and welfare effects, however the increase in magnitudes are not too large.

5.1 Welfare comparisons

In order to measure welfare, a utilitarian social welfare function is employed where each household is weighed equally. The social planner is assumed to assign equal value to the utility from consumption of all members of a household at a given time. The utility in the future is discounted by the discount factor β of the household. Hence, the social welfare in a balanced growth path equilibrium with output growth rate g is given by

$$\begin{aligned} W &= \sum_{t=0}^{\infty} \beta^t \int_{m=0}^{1} \left(\frac{c_{c,m,t}^{1-\omega}}{1-\omega} + \frac{c_{y,m,t-1}^{1-\omega}}{1-\omega} + \frac{c_{o,m,t-2}^{1-\omega}}{1-\omega} \right) dm \\ &= \frac{\int_{m=0}^{1} \left(c_{c,m,0}^{1-\omega} + c_{y,m,-1}^{1-\omega} + c_{o,m,-2}^{1-\omega} \right) dm}{(1-\omega)(1-\beta(1+g)^{1-\omega})} \end{aligned} \tag{21}$$

The welfare comparisons between different economies will be conducted by comparing the balanced growth path equilibria.[32] In order to make two different economies A and B comparable, both economies will be started at the same aggregate productivity level $\bar{z}_0^A = \bar{z}_0^B = 1$. Let $\nu > 0$ be the scalar such that multiplying every agent's consumption in economy A with ν results in a welfare number equivalent to the one in economy B. Simple algebra reveals that ν is given by

$$\nu = (W^B/W^A)^{1/(1-\omega)} \tag{22}$$

where W^A and W^B denote the welfare in economies A and B respectively. The welfare gain or loss a move from economy A to economy B provides in consumption equivalent terms is given by $\nu - 1$. This welfare measure is used in all quantitative exercises.

5.2 Shutting down the credentialing spending channel

How does the misallocation of talent affect economic growth and social welfare? In order to address this question, a simple hypothetical thought experiment will be conducted. Recall that individuals

[32] Hence this analysis ignores the welfare effects of the transition to the new steady state.

heterogeneous in innate ability, early childhood education and wealth are able to receive high quality tertiary education if they can achieve a high enough score given by

$$\tilde{s}(l(h,a),n) = (1-\nu)l(h,a) + \nu n + \epsilon_j.$$

The score of an individual is partially influenced by the actual individual productivity $l(h,a)$, partially by the credentialing spending n, and partially by the random shock ϵ_j. Given the scarcity of high quality tertiary education, increasing the growth rate of the economy is only possible through improving the composition of the individuals who get high quality tertiary education in terms of individual productivity. If the influence of credentialing spending could be diminished such that $\nu = 0$, the scores of the individuals would be perfectly correlated with their actual individual productivity sans the random shock. This would result in highly talented individuals ending up in the innovation sector, where they can contribute to the aggregate productivity growth. Following this line of thought, the economy calibrated in Section 4 is taken, and the parameter ν is set to 0. This hypothetical economy is then compared to the baseline economy.

TABLE 9: SHUTTING DOWN THE CREDENTIALING SPENDING CHANNEL

Variable	*Baseline*	$\nu = 0$	Change
Extensive margin, income effect	0.19	0.05	-73.7%
Extensive margin, education effect	0.07	0.15	114%
Intensive margin, income effect	0.08	0.06	-25.0%
Intensive margin, education effect	0.22	0.20	-9.09%
Yearly GDP growth rate	2.00%	2.21%	10.4%
Education spending/GDP	8.55%	10.2%	19.1%
Aggregate skilled labor, L_s	0.48	0.62	28.4%
Aggregate unskilled labor, L_u	1.91	2.00	4.69%
Mean innate ability of skilled workers, a	2.08	2.57	23.4%
Mean early childhood education of skilled workers, h	2.27	2.96	30.1%
Mean parental wealth of skilled workers, y_o	0.87	0.84	-4.32%
Mean bequests received of skilled workers, b	0.49	0.25	-49.5%
Wage income Gini index	0.52	0.56	6.61%
Log 90/10 ratio	1.17	1.20	3.10%
Log 90/50 ratio	0.52	0.57	9.30%
Log 50/10 ratio	0.65	0.64	-1.88%

Table 9 displays the values of several statistics of interest in the baseline and hypothetical economies and how much they change in percentage terms. The first four rows display how the effects of income and education in the extensive and intensive margins change. In the baseline economy, income effect dominated in the determination of the chances of getting an innovation job, whereas education effect dominated in the prediction of productivity conditional on becoming an inventor. Now that the credentialing channel has been shut down, education effect dominates in both the extensive and intensive margins. Thus the people who would perform better as inventors and those who actually become inventors largely coincide.

The annual GDP growth rate changes from 2.00% to 2.21%, a large increase. This is caused by a 28.4% increase in the aggregate skilled labor supply L_s. Investigating the changes in the characteristics of the people who become inventors reveals that this is driven by higher quality individuals in terms of both innate ability and early childhood education. The mean innate ability a of inventors increases by 23.4%, indicating a better allocation of naturally talented individuals to where their contribution would be the greatest. Furthermore, these individuals also receive more early childhood education investment when they are children, further increasing the average individual productivity of inventors.

Looking at the parental backgrounds of the inventors, it is observed that the mean parental wealth is slightly lower by -4.32%. However the mean bequests received fall tremendously by 49.5%. This is driven by two effects working in the same direction: (1) since $\nu = 0$, it is no longer possible for less talented children with wealthier parents to outperform the more talented but less wealthy competitors in score by outspending them in credentialing, (2) given that their children do not need to spend any money on credentialing; the parents do not deem it necessary to leave large bequests, spending some of the extra windfall for their own consumption, and the rest on the productive early childhood education investment which improves individual productivity l and score $\tilde{s}$ simultaneously.

The inequality measures tell a different story: The decrease in the misallocation of talent is beneficial for economic growth, but it also leads to a more unequal society in terms of income. The Gini index increases from 0.52 to 0.57. Examining the income ratios is more revealing: Log 90/10 ratio increases by 3.10%, exhibiting an increase in the gap between the rich and the poor. However log 90/50 ratio increases at a much higher rate of 9.30%, whereas log 50/10 ratio decreases by 1.88%. These results indicate that the increase in inequality is largely driven by the upper tail of the income distribution. As more naturally talented individuals have better chances at becoming inventors, they are also able to earn higher incomes, drifting away from the rest of the workers.

As a combined result of all of these changes, the welfare in the hypothetical economy is 5.93% higher than the baseline economy in consumption equivalent terms. However, it is important to keep in mind that the hypothetical economy is still far away from the first best. Although the misallocation of talent in the tertiary education stage is reduced to the effect of the randomness inherent in the allocation process only, the early childhood education investment in children is still a function of parental wealth. Thus there is still room for improvement. In addition, the egalitarian social welfare function assigns importance into equalizing outcomes between households in terms of consumption, so holding everything constant, there are also potential gains from redistribution of resources. The following subsection discusses a potential government policy which can address a combination of the listed concerns simultaneously.

5.3 Progressive bequest taxation

The previous thought experiment shows that reducing the misallocation of talent in the economy by shutting down the credentialing spending channel can lead to significant gains in growth and welfare. Can a benevolent government achieve similar gains by utilizing available policy options in a decentralized economy? To this end, socially optimal progressive bequest taxes will now considered. In order to reduce the cost of computation, a particular functional form is assumed with the scale parameter τ_s and the progressivity parameter τ_p such that the budget constraint of the old adults in the decision problem given in (12) becomes

$$c_o + \left(\frac{b}{1-\tau_s}\right)^{\frac{1}{1-\tau_p}} \leq y_o$$

which is equivalent to the old budget constraint if $\tau_s = \tau_p = 0$. All the collected taxes are then transferred to the young adults as a type-independent lump sum transfer Tr, changing the equation that determines y_y in (13) to

$$y_y \quad = \quad \left(w_{j_y} + \frac{w'_{j_y}}{1+r'}\right) l_y(h_y, a_y) + b - c_n(n) + Tr.$$

In order to prevent lump sum taxes, $Tr \geq 0$ is imposed, and the government must balance its budget every period.

The welfare maximizing values of τ_s and τ_p are found to be 0.125 and 0.171 respectively. The bequest tax schedule implied by these two values is quite progressive: The average bequest tax rate faced by the top 1% is 12.1%, whereas this number falls to 9.70% for the top 5%, and 4.18% for the top 10%. In fact, when the transfers are also taken into account, the bottom 95% of the households are net recipients, whereas only the top 5% pay into the system. Furthermore, as it will be demonstrated later on, this progressive taxation scheme does not result in a less productive society: the aggregate productivity of the inventors and the growth rate of output are higher in this alternative economy. Hence the increased equity does not come at the cost of reducing efficiency.

Table 10 shows how the statistics of interest change compared to the baseline under the optimal progressive bequest taxation policy. Looking at the regression targets, and the extensive margin in particular, income loses its explanatory power by 10.5% of its value, whereas that of education increases by 14.3%. The effects on the intensive margin are much more pronounced, where income loses 75% of its explanatory power, and education completely dominates. All of these targets point towards a decrease in the misallocation of talent.

Table 10: Optimal progressive bequest taxation results

Variable	*Baseline*	*Optimal b tax*	Change
Extensive margin, income effect	0.19	0.17	-10.5%
Extensive margin, education effect	0.07	0.08	14.3%
Intensive margin, income effect	0.08	0.02	-75.0%
Intensive margin, education effect	0.22	0.27	22.7%
Yearly GDP growth rate	2.00%	2.05%	2.50%
Education spending/GDP	8.55%	9.13%	6.85%
Aggregate skilled labor, L_s	0.48	0.51	6.29%
Aggregate unskilled labor, L_u	1.91	1.93	0.94%
Mean innate ability of skilled workers, a	2.08	2.15	3.33%
Mean early childhood education of skilled workers, h	2.27	2.47	8.90%
Mean parental wealth of skilled workers, y_o	0.87	0.85	-3.05%
Mean bequests received of skilled workers, b	0.49	0.43	-10.6%
Wage income Gini index	0.52	0.53	1.92%
Log 90/10 ratio	1.17	1.17	0.54%
Log 90/50 ratio	0.52	0.52	0.00%
Log 50/10 ratio	0.65	0.66	0.01%

The growth rate of the economy increases to 2.05% from its baseline value of 2.00%, which corresponds to one quarter of the effect observed in the case of $\nu = 0$. This is caused by the increase in the aggregate skilled labor supply L_s by 6.29%. Examining the mean innate ability a and early childhood education h of inventors, the increase of quality in the composition is driven more by early childhood education (8.90%) rather than innate ability (3.33%). So it can be argued that the optimal bequest taxes contribute to the growth rate of the economy more through reducing the suboptimal investment in early childhood education rather than allocating higher innate ability people to the innovation sector. However, both channels have a positive contribution regardless of their relative power.

In contrast to the thought experiment where credentialing spending is shut down, the increase in the growth rate of the economy is not accompanied by a significant increase in income inequality. The inequality metrics under the optimal taxation policy have very similar values to their baseline values. This is caused by the redistributive nature of the optimal tax policy. As a result of this, even though the growth gain is one quarter of the $\nu = 0$ case, the welfare gain is calculated to be slightly higher: 6.20% in consumption equivalent terms.

5.4 Recalibration with lower intergenerational earnings persistence

The intergenerational persistence of innate ability ρ is an important parameter of the model, the value of which has an important bearing on quantitative counterfactuals. Since innate ability is not directly observable, the value of ρ is indirectly inferred by trying to match the intergenerational

Table 11: Shutting down the credentialing spending channel - Low earnings persistence

Variable	*Baseline*	$\nu = 0$	Change
Extensive margin, income effect	0.05	0.03	-40.0%
Extensive margin, education effect	0.03	0.05	66.7%
Intensive margin, income effect	0.05	0.07	40.0%
Intensive margin, education effect	0.06	0.09	50.0%
Yearly GDP growth rate	2.00%	2.29%	14.5%
Education spending/GDP	7.09%	10.7%	51.5%
Aggregate skilled labor, L_s	0.37	0.52	42.5%
Aggregate unskilled labor, L_u	1.72	1.92	11.9%
Mean innate ability of skilled workers, a	1.85	2.45	32.5%
Mean early childhood education of skilled workers, h	1.43	2.18	52.7%
Mean parental wealth of skilled workers, y_o	0.70	0.74	5.34%
Mean bequests received of skilled workers, b	0.45	0.28	-38.3%
Wage income Gini index	0.49	0.52	6.37%
Log 90/10 ratio	1.03	1.00	-3.11%
Log 90/50 ratio	0.48	0.46	-3.71%
Log 50/10 ratio	0.56	0.54	-2.59%

correlation of earnings (IGE) generated in the model with that found in the data. However, the exact value of IGE in the U.S. over the time period is not a settled topic in the literature.[33] Although consistent with the highly persistent effects of income and education discovered in Section 3, the value of 0.70 targeted in the baseline analysis is on the higher end of the estimates found in the literature. This section repeats the calibration exercise in Section 4 with a lower IGE target of 0.45, and assesses its effects.

The calibrated values of most parameters remain very similar to the results in Table 7, with the exception of intergenerational persistence of innate ability, ρ. This falls from 0.70 to 0.40, a very significant decrease. As a result, the effects of income and education on both margins fall, as well as the differences between the effects for a given regression. The earnings inequality in the steady state is also lower.

How does the lower value of ρ effect the counterfactual experiments? In order to answer this question, the hypothetical thought experiment in Section 5.2 is repeated under the new calibration. Table 11 summarizes the results of shutting down credentialing spending by setting $\nu = 0$. The output growth rate of the economy increases from 2.00% to 2.29%, driven by a huge 42.5% increase in aggregate skilled labor supply. Compared to the baseline economy, the welfare gain is found to be 6.63% in consumption equivalent terms.

[33]See the seminal work of Solon (1999) on the issue, and Black and Devereux (2010), Chetty, Hendren, Kline, and Saez (2014) and the references therein for a recent survey of the literature.

These values are slightly higher compared to those found in Section 5.2. Why is this the case? Inspection reveals that this is caused by a higher degree of initial misallocation of talent in the low IGE economy. Under the baseline calibration, due to the higher persistence of innate ability ρ at 0.70, the rich and the talented largely coincide in the stationary equilibrium. When this persistence is lower at 0.40, the chances of a genius being born to a comparatively poor household are higher. As a result of this, the mean innate ability a of inventors is lower before the shutdown of the credentialing channel. Hence, the growth and welfare implications are amplified when ρ is lower.

6 Conclusions

This paper develops a model of misallocation of talent in the innovation sector. Workers in the economy are finitely-lived, and heterogeneous in terms of wealth, early childhood education, and innate ability. The sectors in the economy are separated into production and innovation, where the latter serves to improve the productivity of the prior. The training necessary to become a worker in the innovation sector is scarce. Agents compete against each other in order to acquire this scarce training so that they can get innovation sector jobs that pay more. They use productive early childhood education investment as well as (socially) unproductive credentialing spending in order to increase their chances. Financial frictions in the form of a non-negative bequest constraint and the inability to insure against idiosyncratic risk, coupled with the misalignment of private and social incentives result in a misallocation of talent across the two sectors. The nature and magnitude of this misallocation of talent are examined.

Empirical analysis makes use of three sets of micro-data—NBER USPTO Utility Patents Grant Database, The Careers and Co-Authorship Networks of U.S. Patent-Holders, and IPUMS-USA 1930 5% Sample—that were previously unlinked in order to establish two new stylized facts: (1) People from richer backgrounds are more likely to become inventors; but those from more educated backgrounds are not. (2) People from more educated backgrounds become more prolific inventors; but those from richer backgrounds exhibit no such aptitude. This discrepancy suggests a misallocation of talent in the innovation sector, which motivates the development of a model that can generate the correlation patterns observed in the data. The results are robust to the use of alternative patent and inventor quality measures, as well as potential biases that might be caused by immigration and similar demographic changes.

The developed model is calibrated to match data targets including aggregate moments of the U.S. economy such as the yearly long-run output growth and real interest rates and the ratio of education spending to GDP; moments obtained using micro data such as intergenerational correlation of earnings and wealth and various inequality measures regarding the earnings distribution; as well as data targets taken from the original empirical analysis such as the effect of income and education on the probability of getting an innovation sector job, and the productivity conditional on having one. The calibrated model is then used to explore how the misallocation of talent between the production and innovation sectors is generated, and the findings suggest that the welfare effects of this misallocation might be substantial.

The quantitative analysis reveals that if the credentialing spending channel could be shut down, the aggregate output growth rate would increase from 2.00% to 2.21%, leading to a welfare gain of 5.93% in consumption equivalent terms. Another quantitative experiment that seeks to calculate the socially optimal bequest taxation policy reveals that the growth rate could be increased to 2.05% even in a decentralized market economy by leveling the playing field and reducing the effect of suboptimal early childhood education spending due to financial frictions. The resulting welfare gain is quite significant at 6.20%. A robustness analysis is conducted to show how the model performs when different calibration targets are chosen, and the quantitative results remain largely similar.

The stylized facts established in the empirical analysis are quite provoking, and the model suggests that reducing the existing misallocation of talent in the economy might yield significant welfare gains through an increase in the long-run output growth rate. Given how important the upper tail of the talent distribution is in generating the ideas that drive economic progress, it is likely that policies that alleviate the misallocation through reducing wealth inequality or financial frictions might be desirable. Future research is needed to establish more detailed policy responses that take additional life-cycle elements into account, as well as the welfare implications of the transition to the new steady-state. The empirical methodology used in the paper can also be applied in any other sector where surname-level information is available, which would considerably expand our understanding of the allocation of talent in other sectors, as well as the intergenerational dynamics of socioeconomic status.

References

Acemoglu, D. (2009): *Introduction to Modern Economic Growth*. Princeton University Press.

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

Aghion, P., U. Akcigit, A. Bergeaud, R. Blundell, and D. Hemous (2015): "Innovation and Top Income Inequality," INSEAD Working Paper No. 2015/50/EPS.

Aghion, P., U. Akcigit, and P. Howitt (2013): "What Do We Learn From Schumpeterian Growth Theory?," National Bureau of Economic Research WP 18824.

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

——— (2009): *The Economics of Growth*. MIT Press.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Aiyagari, S. R., J. Greenwood, and A. Seshadri (2002): "Efficient Investment in Children," *Journal of Economic Theory*, 102(2), 290–321.

Akcigit, U., M. A. Celik, and J. Greenwood (2015): "Buy, Keep, or Sell: Economic Growth and the Market for Ideas," Econometrica, forthcoming.

Alvarez, F. E., F. J. Buera, and R. E. Lucas (2013): "Idea Flows, Economic Growth, and Trade," National Bureau of Economic Research WP 19667.

Anderberg, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," 93(9-10), 1017–1026.

Banerjee, A. V., and A. F. Newman (1993): "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101(2), 274–298.

Becker, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.

Becker, G. S., and N. Tomes (1979): "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87(6), 1153–1189.

——— (1986): "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4(3), 1–39.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1994): "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment," *Journal of Political Economy*, 102(6), 1131–1174.

Behrman, J. R., P. Taubman, and T. J. Wales (1977): "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success," in *Kinometrics: Determinants of Socioeconomic Success within and between Families*, ed. by P. Taubman. North-Holland.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. van Reenen (2015): "The Lifecycle of Inventors," Working Paper in progress.

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4), 352–365.

Black, S. E., and P. J. Devereux (2010): "Recent Developments in Intergenerational Mobility," National Bureau of Economic Research WP 15889.

Bohacek, R., and M. Kapicka (2009): "Optimal Human Capital Policies," 55(1), 1–16.

Charles, K. K., and E. Hurst (2003): "The Correlation of Wealth Across Generations," *Journal of Political Economy*, 111(6), 1155–1182.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014): "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," National Bureau of Economic Research WP 19843.

Clark, G. (2014): *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press.

Corrado, C., C. Hulten, and D. Sichel (2009): "Intangible Capital and US Economic Growth," *Review of Income and Wealth*, 55(3), 661–685.

Cunha, F., and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97(2), 31–47.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006): "Interpreting the Evidence on Life Cycle Skill Formation," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch. Elsevier.

Dahl, G. B., and L. Lochner (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," *American Economic Review*, 102(5), 1927–1956.

Fernandez, R., and J. Gali (1999): "To Each According to...? Markets, Tournaments, and the Matching Problem with Borrowing Constraints," *Review of Economic Studies*, 66(4), 799–824.

Findeisen, S., and D. Sachs (2013): "Education and Optimal Dynamic Taxation: The Role of Income-contingent Student Loans," University of Konstanz WP.

Galor, O. (2009): *Inequality and Economic Development: The Modern Perspective*. Edward Elgar Publishing.

Galor, O., and J. Zeira (1988): "Income Distribution and Investment in Human Capital: Macroeconomics Implications," .

——— (1994): "Income Distribution and Macroeconomics," *Review of Economic Studies*, 60(1), 35–52.

Grochulski, B., and T. Piskorski (2010): "Risky Human Capital and Deferred Capital Income Taxation," *Journal of Economic Theory*, 145(3), 908–943.

Guner, N., G. Ventura, and Y. Xu (2008): "Macroeconomic Implications of Size-dependent Policies," *Review of Economic Dynamics*, 11(4), 721–744.

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

——— (2005): "Market Value and Patent Citations," *RAND Journal of Economics*, 36(1), 16–38.

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

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2013): "The Allocation of Talent and U.S. Economic Growth," National Bureau of Economic Research WP 18693.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124(4), 1403–1448.

Jones, C. I. (2013): "Misallocation, Economic Growth, and Input-Output Economics," in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, vol. II, pp. 419–456. Cambridge University Press.

Jovanovic, B. (2014): "Misallocation and Growth," *American Economic Review*, 104(4), 1149–1171.

Kapicka, M. (2013): "Optimal Mirrleesean Taxation with Unobservable Human Capital Formation," University of California, Santa Barbara WP.

Kapicka, M., and J. Neira (2013): "Optimal Taxation in a Life-Cycle Economy with Endogeneous Human Capital Formation," University of California, Santa Barbara WP.

Kaplow, L. (2005): "The Value of a Statistical Life and the Coefficient of Relative Risk Aversion," *Journal of Risk and Uncertainty*, 31(1), 23–34.

Knowles, J. (1999): "Parental Decisions and Equilibrium Inequality," Ph.D. thesis, University of Rochester.

Krueger, D., and A. Ludwig (2013): "Optimal Capital and Progressive Labor Income Taxation with Endogenous Schooling Decisions and Intergenerational Transfers," University of Pennsylvania WP.

Lai, R., A. D'Amour, and L. Fleming (2009): "Disambiguation and Co-author Networks of the US Patent Inventor Database," Harvard Business School WP.

Lee, S. Y., and A. Seshadri (2014): "On the intergenerational transmission of economic status," Unpublished manuscript, University of Wisconsin-Madison.

Lucas, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3–42.

——— (2009): "Ideas and Growth," *Economica*, 76(301), 1–19.

Lucas, R. E., and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122(1), 1–51.

Maoz, Y. D., and O. Moav (1999): "Intergenerational Mobility and the Process of Development," *Economic Journal*, 109(458), 677–697.

OECD (2013): *Education at a Glance 2013: OECD Indicators*. OECD Publishing.

Olivetti, C., and M. D. Paserman (2013): "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1930," National Bureau of Economic Research WP 18822.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11(4), 707–720.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), 71–102.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010): "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Harvard Business School WP.

Solon, G. (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, pp. 1761–1800. Elsevier.

Spence, A. M. (1973): "Job Market Signalling," *Quarterly Journal of Economics*, 87(3), 355–374.

Stantcheva, S. (2014): "Optimal Taxation and Human Capital Policies over the Life Cycle," Massachusetts Institute of Technology WP.

Stiglitz, J. E. (1975): "The Theory of "Screening," Education, and the Distribution of Income," *American Economic Review*, 65(3), 283–300.

Word, D. L., C. D. Coleman, R. Nunziata, and R. Kominski (2008): "Demographic Aspects of Surnames from Census 2000," Discussion paper, U.S. Census Bureau.

Appendices

A Theory Appendix

A.1 Proof of Theorem 1:

The static profit maximization of a firm is stated as follows:

$$\Pi(z,\Theta) = \max_{k,l_u\geq 0}\{z^{\zeta}k^{\kappa}l_u^{\lambda} - (r+\delta)k - w_u l_u\} \tag{23}$$

First order conditions imply $l_u^* = \frac{\lambda o^*}{w_u}$ and $k^* = \frac{\kappa o^*}{r+\delta}$, hence we have

$$\begin{aligned} o^* &= z^{\zeta}\left(\frac{\kappa o^*}{r+\delta}\right)^{\kappa}\left(\frac{\lambda o^*}{w_u}\right)^{\lambda} \\ o^* &= \left[\left(\frac{\kappa}{r+\delta}\right)^{\kappa}\left(\frac{\lambda}{w_u}\right)^{\lambda}\right]^{1/\zeta} z \end{aligned} \tag{24}$$

and the profits are simply equal to $\Pi(z,\Theta) = \zeta o^*$. From the unskilled labor market clearing condition, we get

$$\begin{aligned} L_u &= \int l_u^*(z)dZ(z) \\ L_u &= \frac{\lambda}{w_u}\left[\left(\frac{\kappa}{r+\delta}\right)^{\kappa}\left(\frac{\lambda}{w_u}\right)^{\lambda}\right]^{1/\zeta}\int z dZ(z) \\ \left(\frac{w_u}{\lambda}\right)^{\frac{\lambda+\zeta}{\zeta}} &= \left(\frac{\kappa}{r+\delta}\right)^{\kappa/\zeta}\frac{\bar{z}}{L_u} \\ w_u &= \lambda\left(\frac{\kappa}{r+\delta}\right)^{\frac{\kappa}{\lambda+\zeta}} L_u^{-\frac{\zeta}{\lambda+\zeta}}\bar{z}^{\frac{\zeta}{\lambda+\zeta}} \end{aligned} \tag{25}$$

where $\bar{z} \equiv \int z dZ(z)$. Note that since L_u is constant along the balanced growth path, the unskilled wage rate grows with gross rate $(1+g_z)^{\zeta/(\lambda+\zeta)}$. Similarly, the capital market clearing condition yields

$$
\begin{aligned}
K &= \int k^*(z)dZ(z) \\
K &= \frac{\kappa}{r+\delta}\left[\left(\frac{\kappa}{r+\delta}\right)^{\kappa}\left(\frac{\lambda}{w_u}\right)^{\lambda}\right]^{1/\zeta}\int z dZ(z) \\
\left(\frac{r+\delta}{\kappa}\right)^{\frac{\kappa+\zeta}{\zeta}} &= \left(\frac{\lambda}{w_u}\right)^{\lambda/\zeta}\frac{\bar{z}}{K} \\
r &= \kappa\left(\frac{\lambda}{w_u}\right)^{\frac{\lambda}{\kappa+\zeta}} K^{-\frac{\zeta}{\kappa+\zeta}}\bar{z}^{\frac{\zeta}{\kappa+\zeta}} - \delta
\end{aligned} \tag{26}
$$

This time since w_u, K and $\bar{z}$ grow over time at gross rates $(1+g_z)^{\zeta/(\lambda+\zeta)}$, $(1+g_z)^{\zeta/(\lambda+\zeta)}$, and $(1+g_z)$ respectively, the interest rate will be constant along the balanced growth path. Define $\tilde{K} = K/\bar{z}^{\frac{\zeta}{\lambda+\zeta}}$ as relative aggregate capital stock. Combining equations (25) and (26) yields the simplified expressions:

$$
\begin{aligned}
w_u &= \lambda\tilde{K}^{\kappa}L_u^{\lambda-1}\bar{z}^{\frac{\zeta}{\lambda+\zeta}} \\
r+\delta &= \kappa\tilde{K}^{\kappa-1}L_u^{\lambda}
\end{aligned}
$$

Plugging the expressions for the unskilled wage rate and the interest rate into the profits yields

$$
\begin{aligned}
\Pi(z,\Theta) &= \zeta\tilde{K}^{\kappa}L_u^{\lambda}\frac{z}{\bar{z}^{\lambda/(\lambda+\zeta)}} \\
&\equiv \pi\frac{z}{\bar{z}^{\lambda/(\lambda+\zeta)}}
\end{aligned} \tag{27}
$$

where π is a time invariant constant.

Define $\hat{z} \equiv z/\bar{z}^{\lambda/(\lambda+\zeta)}$, $\tilde{z} \equiv \bar{z}^{\zeta/(\lambda+\zeta)}$ and $\tilde{w}_s \equiv w_s/\tilde{z}$. The guess and verify method will be used to solve the value function of the firm in the innovation decision problem. Assume the value function of the firm has the form $V(z,\Theta) = v_1\hat{z} + v_2\tilde{z}$ where v_1 and v_2 are scalars. Plugging the solution into the problem, we get:

$$
\begin{aligned}
V(z,\Theta) &= \max_{l_s\geq 0}\left\{\Pi(z,\Theta) + \frac{\chi l_s^{\xi}}{1+r}V(z+\gamma\bar{z},\Theta') + \frac{(1-\chi l_s^{\xi})}{1+r}V(z,\Theta') - w_s l_s\right\} \\
&= \pi\hat{z} + \frac{v_1\hat{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}} + \frac{v_2(1+g_z)^{\zeta/(\lambda+\zeta)}\tilde{z}}{(1+r)} \\
&\quad + \max_{l_s\geq 0}\left\{\frac{\chi l_s^{\xi}}{1+r}\frac{v_1\gamma}{(1+g_z)^{\lambda/(\zeta+\lambda)}} - \tilde{w}_s l_s\right\}\tilde{z}
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
&= \pi\hat{z} + \frac{v_1\hat{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}} + \frac{v_2(1+g_z)^{\zeta/(\lambda+\zeta)}\tilde{z}}{(1+r)} \\
&\quad + \left(\frac{\xi}{\tilde{w}_s}\right)^{\frac{\xi}{1-\xi}} \left[\frac{\chi\gamma v_1}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}} (1-\xi)\tilde{z} \\
&= v_1\hat{z} + v_2\tilde{z}
\end{aligned}
$$

where

$$
v_1 = \frac{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)} - 1}\pi \tag{29}
$$

$$
v_2 = \frac{(1+r)}{(1+r)-(1+g_z)^{\zeta/(\lambda+\zeta)}} \left[\left(\frac{\xi}{\tilde{w}_s}\right)^{\frac{\xi}{1-\xi}} \left[\frac{\chi\gamma v_1}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}} (1-\xi)\right] \tag{30}
$$

$$
l_s^* = \left[\frac{\xi}{\tilde{w}_s}\frac{\chi\gamma v_1}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}}
$$

It is required to verify that w_s grows at gross rate $(1+g_z)^{\zeta/(\lambda+\zeta)}$. Without loss of generality, assume $\Lambda = 1$. Market clearing for skilled labor requires

$$
\begin{aligned}
L_s &= \int l_s^* dZ(z) \\
L_s &= \left[\frac{\xi}{w_s}\frac{\chi\gamma v_1\tilde{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}}\right]^{\frac{1}{1-\xi}} \\
w_s &= \frac{\xi\chi\gamma v_1\tilde{z}}{(1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}L_s^{1-\xi}}
\end{aligned}
$$

$$
w_s = \frac{\xi\chi\gamma\pi\tilde{z}}{((1+r)(1+g_z)^{\lambda/(\zeta+\lambda)}-1)L_s^{1-\xi}} \tag{31}
$$

proving the statement. Also notice that $L_s = l_s^*$. The aggregate productivity evolves according to

$$
\bar{z}' = \bar{z} + \gamma\chi L_s^{\xi}\bar{z}
$$

$$
\Rightarrow g_z = \Gamma L_s^{\xi} \tag{32}
$$

where $\Gamma \equiv \gamma\chi$. This concludes the proof.

A.2 Computational algorithm

Given closed-form solutions for the firm's maximization problem and the resulting system of non-linear equations in Theorem 1, the following computational algorithm is used to solve for the BGP equilibrium of the model:

1. Create grids for y_o, h, b, y_y, a.

2. Guess initial $V_o(y_o, h, a), W(b, h, a), V_y(y_y, a)$.

3. Guess initial $w_u, w_s, r, g, \bar{s}$.

4. Until convergence in value functions according to the sup-norm is achieved, do:

   (a) Solve:

$$\begin{aligned} V_o(y_o, h_y, a_y, \Theta) &= \max_{c_o, b \geq 0} \{u(c_0) + \alpha W(b, h_y, a_y, \Theta)\} \text{ s.t.} \\ c_o + b &\leq y_0 \end{aligned}$$

   Details: Single variable maximization where $b \in [0, y_o]$. One dimensional interpolation is required for evaluation.

   (b) Solve:

$$\begin{aligned} W(b, h_y, a_y, \Theta) &= \max_{n \geq 0} \{E\left[V_y(y_y, a_y, \Theta)|\cdot\right]\} \text{ s.t.} \\ y_y &= \left(w_{j_y} + \frac{w'_{j_y}}{1 + r'}\right) l_y(h_y, a_y) + b - c_n(n) \\ j_y &\sim F(j; l_y(h_y, a_y), n, \Theta) \end{aligned}$$

   Details: Single variable maximization where $n \in [o, \bar{n}]$, where $\bar{n}$ assures positive y_y in the worst case scenario. One dimensional interpolation is required for evaluation. Normal cumulative distribution function is required for calculations. Expectation is calculated over j realization.

   (c) Solve:

$$\begin{aligned} V_y(y_y, a_y, \Theta) &= \max_{c_y, c_c, h'_y, s \geq 0} \{u(c_y) + \alpha u(c_c) + \\ &\quad \beta E[V_o(y'_o, h'_y, a'_y, \Theta')|\cdot]\} \text{ s.t.} \\ y_y &\geq c_y + c_c + c_h(h'_y) + s \\ y'_o &= (1 + r')s \\ a'_y &\sim g(a_y) \\ \Theta' &= T(\Theta) \end{aligned}$$

   Two variable maximization where $s \in [0, y_y]$, $h \in [0, (y_y/\kappa_h)^{1/xi_h}]$, and resulting c_y, c_c must be positive. Two dimensional interpolation is required for evaluation. Expectation is calculated over a' realization.

5. Simulate to calculate capital, skilled and unskilled labor, and fraction of population in each job. One uniform and one normal draw are required for each household and period.

6. Update $w_u, w_s, r, g, \bar{s}$ using simulation results, and go back to (4) up until they are consistent with the market clearing equations and η.

A.3 Aggregate factor demand equations

For computational purposes, it is useful to characterize aggregate factor demands in terms of only factor prices, and factor prices only in terms of aggregate factor demands. This section derives these algebraically using the equations from Appendix A.

In a stationary equilibrium, the aggregate demand for skilled and unskilled labor, L_s and L_u, and the capital rental rate r are constants. The aggregate demand for capital, K, and the wage rates for skilled and unskilled labor, w_s and w_u, grow at the same rate as aggregate output, in proportion to $\tilde{z} = \bar{z}^{\zeta/(\lambda+\zeta)}$. Define normalized aggregate capital stock, skilled and unskilled wage rates as $\tilde{K} = K/\tilde{z}, \tilde{w}_s = w_s/\tilde{z}$ and $\tilde{w}_u = w_u/\tilde{z}$ respectively. First, notice that by only using the definition for $\tilde{w}_u$ and Equation 24, the following identity for π is obtained:

$$\pi = \zeta \left[\left(\frac{\kappa}{r+\delta} \right)^{\kappa} \left(\frac{\lambda}{\tilde{w}_u} \right)^{\lambda} \right]^{1/\zeta} \tag{33}$$

Then we have:

$$L_u = \left(\frac{\kappa}{r+\delta} \right)^{\kappa/\zeta} \left(\frac{\lambda}{\tilde{w}_u} \right)^{\frac{\lambda+\zeta}{\zeta}} \tag{34}$$

$$\tilde{K} = \left(\frac{\kappa}{r+\delta} \right)^{\frac{\kappa+\zeta}{\zeta}} \left(\frac{\lambda}{\tilde{w}_u} \right)^{\frac{\lambda}{\zeta}} \tag{35}$$

$$L_s = \left[\frac{\xi}{\tilde{w}_s} \frac{\chi\gamma\pi}{((1+r)(1+g_z)^{\lambda/(\zeta+\lambda)} - 1)} \right]^{\frac{1}{1-\xi}} \tag{36}$$

Given these equations, it can be verified that:

$$\pi = \zeta \tilde{K}^{\kappa} L_u^{\lambda} \tag{37}$$

Then we have:

$$\tilde{w}_u = \lambda \tilde{K}^{\kappa} L_u^{\lambda-1} \tag{38}$$

$$r + \delta = \kappa \tilde{K}^{\kappa-1} L_u^{\lambda} \tag{39}$$

$$\tilde{w}_s = \frac{\xi\chi\gamma\pi}{((1+r)(1+g_z)^{\lambda/(\zeta+\lambda)} - 1) L_s^{1-\xi}} \tag{40}$$

B Empirical Appendix

Table 12: Intensive Margin Robustness - Alternative Measures II

	log total patents (renewed once) (1975-2006)	log total patents (renewed twice) (1975-2006)	log total patents (top 5% only) (1975-2006)	log total patents (top 20% only) (1975-2006)
log income (1930)	.037***	.033***	.029***	.033***
	(.009)	(.009)	(.008)	(.009)
log education (1930)	.099***	.099***	.075***	.096***
	(.008)	(.008)	(.008)	(.008)
Obs.	78,438	78,438	81,348	81,348
R^2	0.05	0.05	0.03	0.04

Notes: Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 13: Immigration Robustness (1880-1930) - Extensive Margin

	relative representation (1975-2008)					
	(1)	(2)	(3)	(4)	(5)	(6)
income (1930)	.296***	.304***	.323***	.223***	.307***	.293***
	(.014)	(.014)	(.015)	(.017)	(.017)	(.013)
education (1930)	.004	.008	−.003	.003	.021*	.005
	(.007)	(.009)	(.010)	(.005)	(.012)	(.007)
pop. share(1930)/pop. share(1880)						.053***
						(.032)
Obs.	64,308	48,282	48,289	32,168	32,159	64,308
R^2	0.35	0.33	0.38	0.46	0.30	0.35

Notes: Column 1 repeats the last column of Table 2 for surnames which population share ratio is not missing. Columns 2 and 3 repeat the same regression after dropping the top and bottom 25% of the sample according to population share ratio respectively. Columns 4 and 5 repeat the same exercise for the top and bottom halves of the sample respectively. Column 6 repeats the same regression with the whole sample while introducing the population share ratio linearly as a regressor in addition to income and education. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 14: Immigration Robustness (1880-1930) - Intensive Margin

	log quality wtd. total patents (1975-2006)					
	(1)	(2)	(3)	(4)	(5)	(6)
log income (1930)	.022*	.009	.045***	.106***	−.005	.022*
	(.012)	(.012)	(.014)	(.019)	(.017)	(.012)
log education (1930)	.142***	.138***	.141***	.089***	.140***	.141***
	(.011)	(.011)	(.013)	(.018)	(.011)	(.011)
pop. share(1930)/pop. share(1880)						−.003
						(.005)
Obs.	50,529	37,921	37,897	25,269	25,265	64,308
R^2	0.05	0.05	0.05	0.04	0.04	0.35

Notes: Column 1 repeats the last column of Table 3 for surnames which population share ratio is not missing. Columns 2 and 3 repeat the same regression after dropping the top and bottom 25% of the sample according to population share ratio respectively. Columns 4 and 5 repeat the same exercise for the top and bottom halves of the sample respectively. Column 6 repeats the same regression with the whole sample while introducing the population share ratio linearly as a regressor in addition to income and education. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 15: Extensive Margin - Male Only

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.259***		.259***
	(.008)		(.008)
education (1930)		.024***	.000
		(.004)	(.003)
Obs.	107,613	107,613	107,613
R^2	0.24	0.18	0.24

Notes: Data is obtained exclusively from the males in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 16: Intensive Margin - Male Only

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	.072***		.006
	(.008)		(.009)
log education (1930)		.176***	.173***
		(.009)	(.010)
Obs.	76,265	76,265	76,265
R^2	0.02	0.05	0.05

Notes: Data is obtained exclusively from the males in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 17: Extensive Margin - Female Only

	relative representation (1975-2008)	relative representation (1975-2008)	relative representation (1975-2008)
income (1930)	.075***		.082***
	(.013)		(.015)
education (1930)		.024*	−.015
		(.013)	(.014)
Obs.	67,240	67,240	67,240
R^2	0.16	0.15	0.16

Notes: Data is obtained exclusively from the females in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

Table 18: Intensive Margin - Female Only

	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)	log quality wtd. total patents (1975-2006)
log income (1930)	$-.025$		$-.072^{***}$
	(.016)		(.019)
log education (1930)		$.061^{***}$	$.103^{***}$
		(.016)	(.019)
Obs.	16,117	16,117	16,117
R^2	0.02	0.02	0.02

Notes: Data is obtained exclusively from the females in all samples. Robust standard errors in parentheses. Dominant race fixed effects are included the coefficients of which are suppressed for brevity. All variables are normalized by subtracting the mean and dividing by the standard error. Observations are weighted by the share of the surname in the general population obtained from the U.S. decennial census of population (2000). *, ** and *** denote significance at 10, 5 and 1% levels respectively.

C Quantitative Appendix

C.1 Relaxing the scarce high quality tertiary education assumption

The fraction of high quality tertiary education available in the society η is assumed to be exogenously fixed in the model. This means that only a fraction η of the population can receive the education necessary to produce ideas and become inventors. As a result, this assumption implies that the output growth rate of the economy can only be increased by allocating more productive individuals as inventors rather than increasing the share of inventors in the population. How would the counterfactual exercises look like if this assumption was relaxed?

In order to answer this question, the opposite extreme will be considered. Recall that the score threshold $\bar{s}_t$ was chosen such that

$$\eta = \int_{\bar{s}_t}^{\infty} \tilde{s} d\tilde{S}_t(\tilde{s})$$

held. Consider setting the fraction η free and fixing $\bar{s}_t$ instead. In this alternative specification $\bar{s}_t$ denotes a fixed achievement rating in score. Individuals who go have scores greater than this threshold get high quality tertiary education, and the rest do not. The calibration of this alternative model is trivial: The parameter η which was externally calibrated becomes an additional targeted moment, and $\bar{s}$ becomes a parameter.

Table 19 presents the results of repeating the credentialing spending shut down experiment executed in Section 5.2 under this alternative model specification. The changes are quite significant: Now that η is freely chosen, its value increases from 11.6% to 44.9%. This means nearly half of the population is now allocated to the innovation sector. As a result, the skilled labor supply is quadrupled, and the output growth rate is nearly doubled, increasing from 2.00% to 3.45%. As one would expect, the welfare gain from this increase is also calculated to be huge at 107%.

Naturally, these numbers are not to be taken seriously, since the specification does not impose any additional cost on the society for quadrupling the amount of high quality tertiary education provided. Rather, these should be viewed as the extreme upper bound on the growth and welfare numbers that could be achieved by relaxing the fixed η assumption. This example also serves to illustrate the fact that exogenously fixing η is a conservative assumption in terms of putting a discipline on the growth and welfare numbers produced by the model in the counterfactual experiments.

Table 19: Shutting down the credentialing spending channel - Free η

Variable	*Baseline*	$\nu = 0$	Change
Extensive margin, income effect	0.19	0.11	-42.1%
Extensive margin, education effect	0.07	0.09	28.6%
Intensive margin, income effect	0.08	0.16	100%
Intensive margin, education effect	0.22	0.13	-40.9%
Yearly GDP growth rate	2.00%	3.45%	72.3%
Education spending/GDP	8.55%	11.8%	37.9%
Aggregate skilled labor, L_s	0.48	2.09	332%
Aggregate unskilled labor, L_u	1.91	1.39	-27.3%
Mean innate ability of skilled workers, a	2.08	1.78	-14.5%
Mean early childhood education of skilled workers, h	2.27	3.45	51.9%
Mean parental wealth of skilled workers, y_o	0.87	0.96	9.68%
Mean bequests received of skilled workers, b	0.49	0.46	-5.37%
Wage income Gini index	0.52	0.50	-4.91%
Log 90/10 ratio	1.17	1.82	55.5%
Log 90/50 ratio	0.52	1.17	125%
Log 50/10 ratio	0.65	0.65	-0.06%

September 2015, revised

Buy, Keep or Sell:

Economic Growth and the Market for Ideas

by

Ufuk Akcigit, Murat Alp Celik and Jeremy Greenwood*

***Econometrica*, forthcoming**

Abstract

An endogenous growth model is developed where each period firms invest in researching and developing new ideas. An idea increases a firm's productivity. By how much depends on the technological propinquity between an idea and the firm's line of business. Ideas can be bought and sold on a market for patents. A firm can sell an idea that is not relevant to its business or buy one if it fails to innovate. The developed model is matched up with stylized facts about the market for patents in the U.S. The analysis gauges how efficiency in the patent market affects growth.

Keywords: Growth, Ideas, Innovation, Misallocation, Patents, Patent Agents, Research and Development, Search Frictions, Technological Propinquity

JEL Nos: O31, O41

**Affilations and Acknowledgements:* Akcigit, University of Chicago and NBER; Celik, University of Pennsylvania; Greenwood, University of Pennsylvania and NBER. The authors thank Jess Benhabib, Sam Kortum, Yunan Li, Laura Liu, Carlos Serrano, Stefanie Stantcheva, and Randy Wright for comments. Pengfei Han and David Zarruk provided excellent research assistance. Wes Cohen and Mike Roach kindly supplied the authors with data on licensing from the Carnegie Mellon Survey. Akcigit gratefully acknowledges financial support from the Ewing M. Kauffman Foundation and from the Alfred P. Sloan Foundation. Address all correspondence to Ufuk Akcigit: uakcigit@uchicago.edu.

UNITED STATES PATENT OFFICE.

WALTER HUNT, OF NEW YORK, N. Y., ASSIGNOR TO WM. RICHARDSON AND JNO. RICHARDSON.

DRESS-PIN.

Specification of Letters Patent No. 6,281, dated April 10, 1849.

To all whom it may concern:

Be it known that I, WALTER HUNT, of the city, county, and State of New York, have invented a new and useful Improvement in the Make or Form of Dress-Pins, of which the following is a faithful and accurate description.

The distinguishing features of this invention consist in the construction of a pin made of one piece of wire or metal combining a spring, and clasp or catch, in which catch, the point of said pin is forced and by its own spring securely retained. They may be made of common pin wire, or of the precious metals.

See Figure 1 in the annexed drawings (which are drawn upon a full scale, and in which the same letters refer to similar parts,) which figure presents a side view of said pin, and in which is shown the three distinct mechanical features, viz: the pin A, the coiled spring B, and the catch D, which is made at the extreme end of the wire bar C, extended from B. Fig. 2 is a similar view of a pin with an elliptical coiled spring, the pin being detached from the catch D and thrown open by the spring B. Fig. 3 gives a top view of the same. Fig. 4 is a top view of the spring made in a flat spiral coil. Fig. 5 is a side view of the same.

Any ornamental design may be attached to the bar C, (see Figs. 6, 7 and 8,) which combined with the advantages of the spring and catch, renders it equally ornamental, and at the same time more secure and durable than any other plan of a clasp pin, heretofore in use, there being no joint to break or pivot to wear or get loose as in other plans. Another great advantages unknown in other plans is found in the perfect convenience of inserting these into the dress, without danger of bending the pin, or wounding the fingers, which renders them equally adapted to either ornamental, common dress, or nursery uses. The same principle is applicable to hair-pins.

My claims in the above described invention, for which I desire to secure Letters Patent are confined to the construction of dress-pins, hair-pins, &c., made from one entire piece of wire or metal, (without a joint or hinge, or any additional metal except for ornament,) forming said pin and combining with it in one and the same piece of wire, a coiled or curved spring, and a clasp or catch, constructed substantially as above set forth and described.

WALTER HUNT.

Witnesses:
JOHN M. KNOX,
JNO. R. CHAPIN.

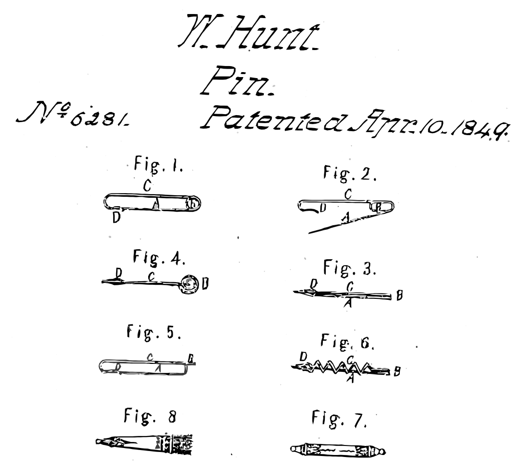


Figure 1: Patent US6281. Walter Hunt's patent for the safety pin.

1 Introduction

New ideas are the seeds for economic growth. The rise in living standards depends on the effectiveness of transforming new ideas into consumer products or production processes. Incarnating an idea into a product or a production process is by no means immediate. Someone must have a vision or an application for the idea and the know-how to implement it. These are often people who work in areas related to the end-use of an idea.

For example, in 1849 Walter Hunt was granted a patent for the safety pin. In the abstract for the patent, Walter Hunt wrote "(t)he distinguishing feature of the invention consist in the construction of a pin made of a piece of wire or metal combining a spring, and a clasp or catch, in which catch the point of the said pin is forced and by its own spring securely retained"–see Figure 1 for his patent application.[1] Hunt was a mechanic by trade and filed patents for various things, such as ice boats, machines for cutting nails, and repeating guns. What is more interesting about this innovation is that Hunt sold his patent to W. R. Grace and Company for about $10,000 (in today's dollars). W. R. Grace and Company mass-produced the safety pin and made millions.

Walter Hunt by no means was an exception. Firms often develop patents that are not close to their primary business activity.[2] Recently released data on the U.S. market for patents indicate that a large fraction of patents are sold by firms, which developed the ideas, to other firms. Specifically, among all the patents registered between 1976 and 2006 in the United States Patent and Trademark Office (USPTO), 16% are traded and this number goes up to 20% among domestic patents.[3] For economic

[1] Patents are publicly disclosed and filed at the United States Patent and Trademark Office. Each patent application has a full description of the invention and drawings to illustrate the embodiments.

[2] Some background material on this is presented in Supplemental Appendix 10.

[3] These numbers do not include patent transfers due to M&A and licensing. They include only firm-to-firm patent transfers and exclude within-firm patent transfers as well as patents sold by individuals. See Appendix 9 for data

progress, not only the possibility of exchange, but also the speed of that process is important. USPTO data shows that new patents are sold among firms on average within 5.48 years (with a standard deviation of 4.58 years).

An analysis of the patent data in Section 3 uncovers some important facts about the nature of these exchanges. A notion of technological propinquity between a patent and a firm is developed. The key findings are:

1. A patent contributes more to a firm's stock market value if it is closer to the firm in terms of technological distance.

2. A patent is more likely to be sold the more distant it is to the inventing firm.

3. A patent is technologically closer to the buying firm than to the selling firm.

The above observations raise important questions that have been left unanswered by the existing literature: How sizeable is the misallocation of ideas across firms? How does efficiency in the market for ideas affect economic growth? Do frictions in the market for ideas lead to more in-house R&D or do they discourage innovations overall? This paper is an attempt to answer these questions.

1.1 The Analysis

To analyze the impact that a market for patents has on the macroeconomy, a search-theoretic growth model is built. The framework is developed in Section 2. Each period firms invest in research and development. Sometimes this process generates an idea, other times it doesn't. Each firm operates within a particular technology class, which is fixed over time. An idea increases a firm's productivity. In the current analysis, the extent to which a firm uses an idea to push forward its productivity depends on the propinquity of the idea to the firm's technology class. A firm may wish to sell an idea that isn't close to its own class. It can do so by using a patent agent. Analogously, the firm might want to purchase an idea through a patent agent if it fails to innovate. Due to search frictions it may take time for a patent agent to find a buyer for a patent. Also, a patent may not be the perfect match for a buyer. R&D by firms leads to growth in the model. Additionally, there is a spillover effect from ideas. A balanced growth path for the model is explicitly characterized. A unique invariant firm-size distribution exists despite the fact that the distribution for productivity across firms is continually fanning out.

The model is calibrated in Section 4 so that it matches certain features of the U.S. aggregate economy, such as the average rate of growth, the long-run interest rate, the share of R&D in GDP, etc. It is also fit to match some facts, presented in Section 3, from the micro data on patents for U.S.

construction.

public firms. Two such facts are descriptive: the share of patents that are sold and the empirical duration distribution. Additionally, some facts from panel-data regression analysis are assembled and targeted using an indirect inference strategy. First, it is shown that a firm's market value is positively related to its stock of patents, controlling for size and other things. Presumably, patents are valuable because they protect important ideas. Second, the closeness of the patents in a firm's portfolio to the firm's line of business matters for market value. Third, the more distant a patent is to a firm's line of business the more likely it is to be sold. Fourth, upon selling a patent there is shrinkage in the distance between the patent and owner's line of business.

Clearly, a market for patents affects the incentive to do R&D. On the one hand, the fact that an idea, which is not so useful for the innovator's own production, can be sold raises the return from engaging in R&D. On the other hand, the fact that a firm can buy an idea reduces the reward from doing R&D. A goal of the analysis is to examine how a patent market affects R&D and, hence, growth. This is done in Section 5.

To gauge the importance of the patent market for economic growth and welfare, a sequence of structured thought experiments is undertaken in Section 5. First, the rate of contact between buyers and sellers in the market is reduced to zero, which is equivalent to shutting down the market. In the absence of the patent market, the equilibrium steady-state growth rate goes from its benchmark value of 2.08% down to 2.02% , resulting in a welfare reduction of 1.18% in consumption equivalent terms. Next, the efficiency of the patent market is successively increased. It is shown that a faster rate of contact between buyers and sellers, where a buyer can find a seller without any delay, increases the growth rate up to 2.46% and leads to a welfare gain of 5.97% relative to the benchmark economy (measured in terms of consumption). In addition, if each seller is matched with the perfect buyer for their patent, then the growth rate increases to 3.05% and a welfare improvement of 14.3% materializes. Last, if the ideas that firms produce are perfectly suited for their own production process (this corresponds to a situation where there is no mismatch between a firm and the idea that it generates) then the growth rate is 3.38%, which results in a welfare gain of 17.8% compared with the baseline model. So, efficiency in the market for patents matters.

Two concerns arise with the focus on patents. First, ideas may be transferred via other means, in particular licensing. The empirical analysis conducted in Section 3 controls for this, to the extent possible. Additionally, the model simulation is redone in Section 6 to allow for ideas to be transferred through licensing as well as patents. The results are not affected in a significant way. Second, perhaps some patents are bought and sold for reasons surrounding litigation. Such sales may have little to do with the transfer of knowledge or increasing productivity. A firm may buy an intrinsically worthless patent to fend off potential litigation, or perhaps to earn profits by threatening litigation (patent trolls). The empirical analysis in Section 3 also attempts to control for this. Additionally, as a robustness check, the model is re-simulated in Section 6 using data from low-litigation sectors. Again, the results appear to be immune to this.

The market for patents is often thought of as being inefficient and illiquid. Buying and selling intellectual property is a difficult activity. Each patent is unique. It may not be readily apparent who the potential buyers and competing sellers even are, especially in situations where enterprises desire to keep their business strategies secret. Buyers and sellers may have very different valuations about the worth of a patent. Patents are often sold through intermediaries. This motivates the search-theoretic framework presented here.

Historically patent agents were often lawyers. Dealing with both patent buyers and sellers, they understood both sides of the market. Inventors used them to file patent applications. So, the lawyers became acquainted with the new technologies that were around. Buyers used them to vet the merits of new technologies. Hence, the lawyers were familiar with the types of patents that were likely to be marketable. This led naturally to the lawyers acting as intermediaries in patent sales. Edward Van Winkle typifies the business. He was a patent agent at the beginning of the 20th century. Van Winkle was a mechanical engineer who acquired a law degree by correspondence course. He was well suited to provide advice on the legal and technical merits of inventions for his clients on both sides of the market. Van Winkle cultivated a network of businessmen, inventors, and other lawyers. Lamoreaux and Sokoloff (2002) detail how he brokered various types of deals with the buyers and sellers of patents. They also document for the period 1870 to 1910 an increased tendency for inventors (especially the more productive ones) to use specialized registered patent agents to handle transactions associated with their patents.

While today's market for patents is sizeable it can be regarded as being thin due to the specialized nature of the knowledge that is embodied in each patent. Thus, the patent market is highly specialized. To date, online intellectual property platforms have failed to arbitrage the market. The sensitivity of intellectual property makes potential buyers and sellers reluctant to reveal information online; they prefer face-to-face dealings with the other party. Also, some buyers may perceive a lemons problem: if the patents were truly valuable, then the sellers should be able to profit by developing the idea themselves or by selling it directly to interested parties.

1.2 Relationship to the Literature

How does the current paper relate to the literature? This is discussed now. On the theory side, the model developed here is in a class of its own, but like all work it is inspired by some important predecessors. The paper contributes to the endogenous growth literature. Ever since Romer's (1986) classic paper, economists have been concerned with how knowledge affects economic growth. The notion that a firm can push forward its productivity by incorporating new ideas in its production process is in Aghion and Howitt (1992). Unlike Aghion and Howitt (1992), this is done here in a competitive environment. The cue for a spillover effect from ideas is in the Romer (1986) growth model.

Recent attention has been directed to developing the micro-foundations of how new ideas spread in an economy. Some work stresses technology diffusion via innovation and imitation [e.g., Jovanovic and MacDonald (1994), and König, Lorenz, and Zilibotti (2012)]. Other research emphasizes matching and other frictions in the transfer of ideas. [See for instance, Benhabib, Perla, and Tonetti (2013), Chiu, Meh, and Wright (2011), Lucas and Moll (2013), and Perla and Tonetti (2013)]. The work here emphasizes matching frictions. It differs from the above papers in a number of significant ways. First, the focus is on an economy where growth is driven by *heterogeneous* ideas that are invented by firms. A firm may not be able to make the best use of the idea it discovers. Second, firms can trade their ideas in a *market* subject to matching frictions. Third, while the growth literature has mainly been theoretical, the current research uses micro data on *patent reassignments* to motivate and discipline the analysis.[4]

The present paper highlights the importance of complementarity (as measured by distance) between the existing knowledge stock of the firm and new patents. These findings naturally relate to work on diversification. In a classic study on diversification and integration, Gort (1962, p 108) states "when faced with a choice among activities that would be equally attractive if they were technologically equidistant from the primary one, a firm will usually undertake those for which technical propinquity to the primary activity is greatest." Gort (1962) provides some early evidence in support of this hypothesis. Figueroa and Serrano (2013) examine the empirical significance of this idea for patenting and licensing activities.

On the empirical side, the data employed here was first used by Serrano (2010, 2015). He uses the fraction of self-citations as a proxy for the fit of an idea to an inventing firm and documents that patents that are not a good fit are more likely to be sold on the market by the inventing firm. A new metric for measuring the distance between ideas and firms is proposed here. Serrano (2010, 2015)'s findings are confirmed. Additionally, new facts on the relationship between a firm's market value and its distance-adjusted patent portfolio are presented. Also, it is shown how the distance between an idea and its owner changes upon sale. The micro data facts that are obtained from the U.S. data are then used here to discipline a search-based endogenous growth model. The model is employed to

[4] Perhaps the closest theoretical work to the current research is by Chiu, Meh, and Wright (2011). Ideas are *homogeneous* in their framework, so there cannot be any misallocation. They are produced by inventors who cannot commercialize them, so *all* ideas are sold. Firms cannot do R&D, hence they must purchase an idea to produce. There are search frictions in their setup: an inventor must find a entrepreneur in order to sell his idea. The work emphasizes financial frictions. In particular, an entrepreneur must have cash on hand to buy an idea. Last, no quantitative work is done.

quantify the misallocation of ideas in the U.S. economy and the contribution of the patent market to economic growth.

The focus on mismatch in ideas connects with recent work on misallocation [see for instance, Acemoglu, Akcigit, Bloom, and Kerr (2013), Guner, Ventura, and Xu (2008), Hsieh and Klenow (2009), and Restuccia and Rogerson (2008)]. That literature has mainly focused on factor misallocations, particularly the allocation of capital and labor across establishments. The current work complements this literature by focusing on differences in total factor productivity that may arise due to a misallocation of ideas, which are a direct ingredient in productivity. Ideas are not necessarily born to their best users. The existence of a market for ideas and its efficiency can have a major impact on mitigating any initial misallocation. Thus, the presence of a market for ideas may contribute significantly to productivity growth. Addressing this question is the focus of the current paper.

2 Model

The theoretical model with perfectly competitive firms is introduced now. The goal is to focus on the potential misallocation of ideas and its consequences for growth and welfare; therefore, the model abstracts from monopoly distortions. Another interesting feature of this setting is that patents serve a new role in this economy: the possibility for trading ideas. Some ideas are better than others for a firm. In the analysis there are two types of ideas: to wit, d-type and n-type. The worth of a d-type idea depends on the distance of the idea to a firm's main line of business. The closer the idea, the more valuable it is. The worth of an n-type idea is unrelated to the distance between the idea and the firm's line of business. To obtain a d-type idea a firm must invest resources, either through R&D or by buying a patent on the market. By contrast, a firm may discover an n-type idea through serendipity for free. The productivity of both types of ideas depends upon the general pool of knowledge in the economy; that is, through osmosis some component of ideas become part of the ether in technology space.[5]

2.1 Environment

Consider an economy, where time flows discretely, with a continuum of firms of unit measure. The firms produce a homogeneous final good using capital and labor. Each firm belongs permanently to some technology class j that resides on a circle with radius $1/\pi$. At each point on the technology circle there are firms of density $1/2$. A firm enters the period with a level of productivity z. At the

[5] A simplified version of the model that connects in a more elementary manner the efficiency of the patent market, and the propinquity of an idea with the firm's line of business, to economic growth is presented in Supplemental Appendix 12.

beginning of a period each firm develops a d-type idea with an *endogenous* probability i. The d-type innovation will be patented and belongs to some technology class k on the circle. The distance between the firm's own technology class, j, and the innovation, k, is denoted by $d(j,k)$. This represents the length of the shortest arc between j and k. Transform this distance measure into a measure of technological propinquity, $x = 1 - d(j,k)$, defined on $[0,1]$. A high value for x indicates that the innovation is close to the firm's technology class. The firm will keep or sell the d-type patent depending on the value for x. The higher x is, the bigger will be the boost to the z, if the firm decides to keep the idea. The value of x is drawn from the distribution function $X(x)$. The technology circle is illustrated in Figure 2. Just before production begins, an n-type idea arrives with an *exogenous* probability $\mathfrak{p}$. The worth of an n-type idea is unrelated to a firm's technology class. In a symmetric equilibrium, at each point on the circle the distribution of firms is the same. The analysis will focus on symmetric equilibrium around this circle. Analyzing one point on the circle is the same as analyzing any other, so there is no need to carry around a location index.

Firms produce output, o, at the end of a period according to the production process

$$o = (e'z')^{\zeta} k^{\kappa} l^{\lambda}, \text{ with } \zeta + \kappa + \lambda = 1, \tag{1}$$

where k and l are the amounts of capital and labor used in production and z' is its end-of-period productivity. The variable e' is a firm-specific idiosyncratic production shock. It is drawn at the end of each period from a log-normal distribution with $E[e'] = 1$ and a standard deviation represented by $\text{STD}(\ln e')$.[6] Labor is hired at the wage rate w. There is one unit of labor available in the economy. Capital is hired at the rental rate $\widetilde{r}$. Observe that there are diminishing returns in capital and labor. Hence, there are profits from producing. These rents are increasing in the firm's productivity, z'. This provides an incentive to do R&D to improve z'. The exponent ζ on the $e'z'$ is an innocuous normalization that results in profits being linear in $e'z'$, as is shown below.

A firm's end-of-period productivity, z', evolves according to the law of motion

$$z' = L(z, x, b; \mathbf{z}) = z + \gamma_d x \mathbf{z} + \gamma_n b \mathbf{z}. \tag{2}$$

Here z is the firm's initial productivity level. The second term gives the increment to productivity from obtaining a d-type patent, where x is the technological propinquity of the patent to the firm and $\mathbf{z}$ is mean of the productivity distribution in the economy at the beginning of the period. The closer a d-type innovation is to a firm's own technology class, as represented by a larger x, the bigger will be the increase in productivity, $\gamma_d x \mathbf{z}$. The third term gives the gain in productivity from acquiring an n-type idea, where $b \in \{0, 1\}$. The expected value of b is given by $E[b] = \mathfrak{p}$. Once an idea is

[6]The e' shock implies that employment, l, will not be a perfect predicator of a firm's market value. This property is important for the quantitative analysis and is discussed in Section 4.

blended into a firm's production process, within the firm's permanent technology class, it loses its individual identity. This assumption implies that there is no need to keep track of a firm's portfolio of patents, which would vary by technology class and age; doing so would be an insurmountable task.

The higher is the economy-wide baseline level of productivity, $\mathbf{z}$, the more valuable a patent is for increasing productivity. This is true for both d-type and n-type patents. Note that $\mathbf{z}$ introduces the usual Romer (1986) type knowledge externality in this economy. Although not modeled formally, this could be because either some forms of knowledge can only be imperfectly protected or because the patents protecting them have expired so that the knowledge formerly embodied in the patents is now freely available for all. Since n-type patents arrive with exogenous probability $\mathfrak{p}$ the firm will benefit from spillovers in a probabilistic sense, even if the firm does not invest any resources in R&D. This is not true for d-type patents, as will be seen. Later, the notation $L(z, 0, b; \mathbf{z})$ will be used to signify the situation where the firm's productivity is not incremented by a d-type innovation in the current period, which is equivalent to setting $x = 0$. One might think that firm would try to discover a d-type idea that is close to its line of business. As was mentioned, the propinquity of a d-type idea to the firm, x, is drawn from the distribution $X(x)$. In the quantitative analysis this will be taken to be the empirical distribution. Hence, the concordance of ideas with their inventors will be the same as in the data. It turns out that $\mathbf{z}$ is also the aggregate state variable in this economy, a fact shown later. That is, only the mean of the distribution for the z's across firms and the evolution of this mean over time matter for the analysis. Assume that $\mathbf{z}$ evolves according to the deterministic aggregate law of motion

$$\mathbf{z}' = \mathbf{T}(\mathbf{z}). \tag{3}$$

Now, at the beginning of a period, firms pick the probability of discovering a d-type idea, i. They do this according to the convex cost function

$$C(i; \mathbf{z}) = \chi \mathbf{z}^{\zeta/(\zeta+\lambda)} i^{1+\rho}/(1+\rho). \tag{4}$$

Cost rises in lock-step fashion with average productivity, $\mathbf{z}$, in the economy. It will be established later that wages, w, are proportional to $\mathbf{z}$ and grow along a balanced growth path at the same rate as $\mathbf{z}^{\zeta/(\zeta+\lambda)}$. As will be seen, this ensures that along a balanced growth path the ratio of aggregate R&D expenditures to GDP remains constant. Aggregate productivity will be a function of the aggregate state of the world represented by $\mathbf{z}$. A firm that successfully innovates can either keep or sell its idea to a patent agent. A firm that does not innovate can try to buy a patent from an agent. A patent on the market survives over time with probability σ. In the analysis σ will be set so that patents have the same expected life as in the U.S. data.[7] But, by letting a patent die stochastically in this fashion,

[7] In the simplified model in Supplemental Appendix 12 patents are allowed to live forever. This does not change things in a material way.

instead of deterministically, there is no need to keep track of a patent's age, a huge simplification.

d-type ideas can be bought and sold on a patent market. A firm that fails to come up with a d-type idea can try to buy one from a patent agent. Innovators are not allowed to buy patents. A firm that draws a d-type idea may sell the associated patent to a patent agent at the price q. This price is determined on a competitive market. Once a patent is sold to an agent the seller cannot use it in the future. A patent agent can only handle one d-type idea at a time. The introduction of patent agents simplifies the analysis. Without this construct the analysis would have to keep track of the portfolio of patents that each firm has for sale. This technical construct is imposed without apology, as in the real world many patents are sold through agents, as was discussed.

Let n_a and n_b represent the numbers of agents and buyers in the market for d-type patents. The total number of matches in the market is given by the matching function

$$M(n_a, n_b) = \eta n_a^{\mu} n_b^{1-\mu}.$$

The matches are completely random. Thus, the odds that an agent will find a buyer are given by

$$m_a(\frac{n_a}{n_b}) = \frac{M(n_a, n_b)}{n_a} = \eta(\frac{n_b}{n_a})^{1-\mu},$$

and similarly that a buyer will find an agent by

$$m_b(\frac{n_a}{n_b}) = \frac{M(n_a, n_b)}{n_b} = \eta(\frac{n_a}{n_b})^{\mu}.$$

This search friction could reflect many things: the hardship of matching buyers and sellers in a thin market for a complicated product or the difficulty of a buyer assessing the quality of a patent for his line of business, inter alia.

The ratio of potential sellers to buyers, n_a/n_b, reflects the slackness of the market. Since agents and buyers are matched randomly, the propinquity between the buyer's technology class and the class of the d-type patent being sold is a random variable. A buyer will incorporate a d-type patent that he purchases into his production process in accordance with the above law of motion for z. The price of the d-type patent is determined by Nash bargaining between the agent and buyer. Represent this price by $p = P(z, x; \mathbf{z})$. The negotiated price will depend on the propinquity of the patent, x, and the state of the buyer's technology, z. The bargaining power of the agent is given by ω. In contrast, the price at which a firm sells its d-type patent to an agent is fixed at q, because the agent doesn't know who he will sell the patent to in the future. The timing of events in the market for d-type patents is portrayed in the right panel of Figure 2. Last, after the d-type patent market closes, an n-type idea may arrive to a firm. For the moment assume that n-type ideas are not traded. A market for n-type ideas is appended onto the model in Section 3.4.

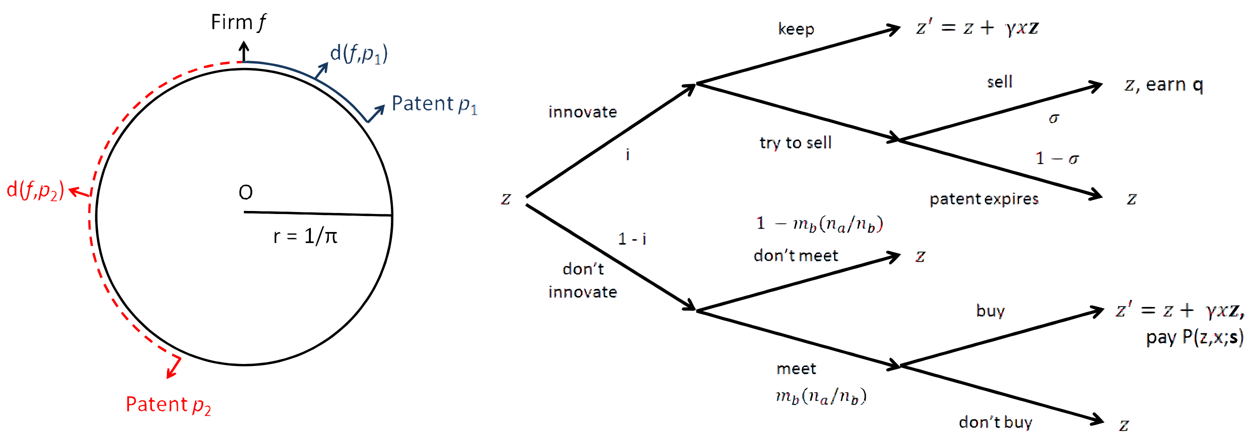


Figure 2: The technology circle (left panel) and the timing of events (right panel) for d-type ideas. Note that n-type ideas arrive after the market for d-type patents closes.

2.2 The Representative Consumer/Worker

In the background of the analysis is a representative consumer/worker. This individual supplies one unit of labor inelastically. The person owns all of the firms in the economy. He also rents out the capital used by firms. Thus, he will earn income from wages, profits and rentals. Capital depreciates at the rate δ. The real return earned by renting capital is $1/r$. (I.e., r is the reciprocal of the gross interest rate. It will play the role of the discount factor in the Bellman equations formulated below.) The individual is assumed to have a momentary utility function of the form $U(c) = c^{1-\varepsilon}/(1-\varepsilon)$, where c is his consumption in the current period and ε is the coefficient of relative risk aversion. He discounts the future at rate β. Last, the representative consumer/worker's goal in life is to maximize his discounted lifetime utility. Since this problem is entirely standard it is not presented.

2.3 Firms

A firm hires labor, l, at the wage rate w, and capital, k, at the rental rate, $\widetilde{r} \equiv 1/r - 1 + \delta$, to maximize profits. It does this at the end of each period after seeing the realized values for e' and z'. Thus, its decision problem is

$$e'\Pi(z';\mathbf{z}) = \max_{k,l}[(e'z')^{\zeta}k^{\kappa}l^{\lambda} - \widetilde{r}k - wl].$$

where $e'\Pi(z';\mathbf{z})$ is the profit function associated with the maximization problem; the fact that this function is multiplicative in e' is established momentarily. The first-order conditions to this maximization problem imply that

$$k = \kappa\frac{o}{\widetilde{r}}, \tag{5}$$

and

$$l = \lambda \frac{o}{w}. \tag{6}$$

Using (1), (5) and (6), it follows that profits are given by

$$e'\Pi(z';\mathbf{z}) = (1-\kappa-\lambda)o = e'z'(1-\kappa-\lambda)[(\frac{\kappa}{\widetilde{r}})^{\kappa}(\frac{\lambda}{w})^{\lambda}]^{1/\zeta}. \tag{7}$$

Again, in equilibrium the rental and wage rates, $\widetilde{r}$ and w, will be functions of the aggregate state of the world. Note that profits are increasing in z' when there are diminishing returns to scale $(1-\kappa-\lambda<1)$. This provides an incentive to innovative.

Let $V(z;\mathbf{z})$ represent the expected present-value of a firm that currently has productivity z and is about to learn whether or not it has come up with a d-type idea. Due to the focus on symmetric equilibrium there is no need ever to record the firm's location on the technology circle. Now, suppose that the firm does not innovate. Then, it will try to buy a d-type patent. With probability $1-m_b(n_a/n_b)$ it will fail to find a patent agent. In that circumstance, the firm's productivity will remain at z; this is equivalent to setting $x=0$ in (2). It may still acquire an n-type patent before the start of production, though, which would allow productivity to increase by $\gamma_n\mathbf{z}$. The expected value of the firm, before the n-type patent shock, is $E[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})]+rE[V(L(z,0,b;\mathbf{z});\mathbf{z}')]$–recall that $b\in\{0,1\}$ is a random variable connected with the n-type idea that takes the value one with probability $\mathfrak{p}$ and that $E[e']=1$.

With probability $m_b(n_a/n_b)$ the firm will meet an agent selling a d-type patent with propinquity x. Two things can happen here: either the firm buys a d-type patent from the agent or it doesn't. The d-type patent sells at the price $p=P(z,x;\mathbf{z})$, which is a function of the buyer's type, z, as well the propinquity of the patent to the firm's technology class, x. The determination of the d-type patent price is discussed below. The firm will only buy the patent if it yields a higher payoff than what it will obtain if it doesn't buy it. If the firm buys a patent its productivity will rise to $z+\gamma_d x\mathbf{z}$. Again, before production begins the firm may also obtain an n-type patent, which would result in a further increase in productivity. The firm's expected value (before the n-type patent shock) will then move up to $E[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})]-P(z,x;\mathbf{z})+rE[V(L(z,x,b;\mathbf{z});\mathbf{z}')]$. If it doesn't buy a d-type patent then its productivity will remain at z. The expected value of the firm will then be $E[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})]+rE[V(L(z,0,b;\mathbf{z});\mathbf{z}')]$. Denote the distribution, over propinquity, for the patent agents by $D(x)$.

The expected discounted present value of the buyer, $B(z;\mathbf{z})$, is easily seen to be

$$\begin{aligned} B(z;\mathbf{z}) = m_b\left(\frac{n_a}{n_b}\right)\int &\{I_a(z,x;\mathbf{z})\left\{E[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})]-P(z,x;\mathbf{z})+rE[V(L(z,x,b;\mathbf{z});\mathbf{z}')]\right\} \\ &+[1-I_a(z,x;\mathbf{z})]\left\{E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right]+rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right]\right\}\}dD(x) \\ &+[1-m_b(\frac{n_a}{n_b})]\left\{E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right]+rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right]\right\}, \end{aligned} \tag{8}$$

where $\mathbf{z}$ evolves according to (3) and

$$I_a(z,x;\mathbf{z}) = \begin{cases} 1 \text{ (sale)}, & \text{if the buyer purchases a patent,} \\ 0 \text{ (no sale)}, & \text{otherwise.} \end{cases} \tag{9}$$

The indicator function $I_a(z,x;\mathbf{z})$, defined above, specifies whether or not the non-innovating firm will buy a d-type patent. The determination of this function is discussed below.

Turn now to the situation where the firm successfully innovates. If it decides to keep the d-type patent then the firm's productivity will be $z+\gamma_d x\mathbf{z}$ as in (2). Productivity may still increase if the firm draws an n-type idea. Before the realization of the n-type patent shock, the firm will have the expected value $K(z+\gamma_d x\mathbf{z};\mathbf{z})$, as given by

$$K(z+\gamma_d x\mathbf{z};\mathbf{z}) = E\left[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})\right] + rE\left[V(L(z,x,b;\mathbf{z});\mathbf{z}')\right], \tag{10}$$

where again $\mathbf{z}$ evolves according to (3) and $b \in \{0,1\}$ is a random variable. Alternatively, it can sell the d-type patent to an agent. Then, its productivity will remain at z (unless it subsequently draws an n-type idea). The value of a seller, $S(z;\mathbf{z})$, is

$$S(z;\mathbf{z}) = E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right] + \sigma q + rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right]. \tag{11}$$

Once the seller puts a d-type patent up for sale at the beginning of the period it expires with probability $1-\sigma$. A firm that innovates will either keep or sell its d-type patent depending on which option yields the highest value. Given this, it is easy to see that the decision to keep or to sell a patent can be formulates as

$$I_k(z,x;\mathbf{z}) = \begin{cases} 1 \text{ (keep)}, & \text{if } K(z+\gamma_d x\mathbf{z};\mathbf{z}) > S(z;\mathbf{z}), \\ 0 \text{ (sell)}, & \text{otherwise.} \end{cases} \tag{12}$$

2.3.1 The Decision to Innovate

The firm's decision to innovate is now cast. With probability i the firm discovers a d-type idea and with probability $1-i$ it doesn't. The firm chooses the probability of discovering a d-type idea subject to the convex cost function $C(i;\mathbf{z})$. Hence, write the innovation decision as

$$V(z;\mathbf{z}) = \max_i \left\{ i \int \{I_k(z,x;\mathbf{z})K(z+\gamma_d x\mathbf{z};\mathbf{z}) + [1-I_k(z,x;\mathbf{z})]S(z;\mathbf{z})\}dX(x) + (1-i)B(z;\mathbf{z}) - C(i;\mathbf{z}) \right\}. \tag{13}$$

The first-order condition associated with this problem is

$$\int \{I_k(z,x;\mathbf{z})K(z+\gamma_d x\mathbf{z};\mathbf{z}) + [1-I_k(z,x;\mathbf{z})]S(z;\mathbf{z})\}dX(x) - B(z;\mathbf{z}) = C_1(i;\mathbf{z}),$$

(where C_1 is the derivative of C with respect to i) so that

$$\begin{aligned} i &= R(z;\mathbf{z}) \qquad (14) \\ &= C_1^{-1}\left(\int \{I_k(z,x;\mathbf{z})K(z+\gamma_d x\mathbf{z};\mathbf{z}) + [1-I_k(z,x;\mathbf{z})]S(z;\mathbf{z})\}dX(x) - B(z;\mathbf{z});\mathbf{z} \right). \end{aligned}$$

2.4 Patent Agents

Turn now to the problem of a patent agent. It buys a d-type idea at the competitively determined price q. With probability $m_a(n_a/n_b)$ it will meet a potential buyer on the market and with probability $1-m_a(n_a/n_b)$ it won't. Denote the distribution of buyers by $G(z)$. The value for an agent, A, with a patent is thus given by

$$A(\mathbf{z}) = m_a\left(\frac{n_a}{n_b}\right)\int\int\left\{I_a(z,x;\mathbf{z})P(z,x;\mathbf{z}) + [1-I_a(z,x;\mathbf{z})]r\sigma A(\mathbf{z}')\right\}dG(z)dD(x)$$
$$+[1-m_a\left(\frac{n_a}{n_b}\right)]r\sigma A(\mathbf{z}'), \tag{15}$$

where $I_a(z,x;\mathbf{z})$ is specified by (9) and is defined formally shortly below. The price of a d-type patent is determined via Nash bargaining. Specifically, p is determined in accordance with

$$\max_p\{E\left[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})\right] - p + rE\left[V(L(z,x,b;\mathbf{z});\mathbf{z}')\right]$$
$$-E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right] - rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right]\}^{1-\omega}$$
$$\times\left[p - r\sigma A(\mathbf{z}')\right]^{\omega}.$$

The first term in braces gives the buyer's surplus. This gives the difference between the value of the firm when it secures a d-type patent and the value when it does not. The second term details the seller's surplus. In standard fashion,

$$p = P(z,x;\mathbf{z}) = \omega\{E\left[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})\right] + rE\left[V(L(z,x,b;\mathbf{z});\mathbf{z}')\right] - E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right]$$
$$-rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right]\} + (1-\omega)r\sigma A(\mathbf{z}'), \tag{16}$$

whenever both the buyer's and seller's surpluses are positive. The price lies between $r\sigma A(\mathbf{z}')$ and $E\left[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})\right] + rE\left[V(L(z,x,b;\mathbf{z});\mathbf{z}')\right] - E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right] - rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right]$; if the former is above the latter then no solution exists. Now, define $I_a(z,x;\mathbf{z})$ in the following manner:

$$I_a(z,x;\mathbf{z}) = \begin{cases} 1, \quad \text{if} & r\sigma A(\mathbf{z}') \leq p \leq E\left[\Pi(L(z,x,b;\mathbf{z});\mathbf{z})\right] + rE\left[V(L(z,x,b;\mathbf{z});\mathbf{z}')\right] \\ & -E\left[\Pi(L(z,0,b;\mathbf{z});\mathbf{z})\right] - rE\left[V(L(z,0,b;\mathbf{z});\mathbf{z}')\right], \\ 0, \quad \text{otherwise.} & \end{cases} \tag{17}$$

2.5 Symmetric Equilibrium Along a Balanced Growth Path

The focus of the analysis is solely on symmetric equilibrium along a balanced growth path. A formal analysis of the model's balanced growth path is contained in Theory Appendix 8. Before starting, define the aggregate level of productivity, $\mathbf{z}$, its gross rate of growth, $\mathbf{g}$, and the aggregate level of innovation, $\mathbf{i}$, by

$$\mathbf{z} \equiv \int zdZ(z),\ \mathbf{g} \equiv \frac{\int z'dZ'(z')}{\int zdZ(z)},\ \text{and}\ \mathbf{i} \equiv \int R(z;\mathbf{z})dZ(z). \tag{18}$$

In equilibrium the demand for labor must equal the supply of labor. Recall that there is one unit of labor in the economy. Let $Z'(z')$ represent the end-of-period distribution of z' across firms. Now, using (1), (5) and (6) it is easy to deduce that the labor, l, demanded by a firm is given by

$$l = \left(\frac{\kappa}{\widetilde{\widetilde{r}}}\right)^{\kappa/\zeta} \left(\frac{\lambda}{w}\right)^{(\zeta+\lambda)/\zeta} e'z'. \tag{19}$$

Equilibrium in the labor market then implies that

$$\int \left(\frac{\kappa}{\widetilde{\widetilde{r}}}\right)^{\kappa/\zeta} \left(\frac{\lambda}{w}\right)^{(\zeta+\lambda)/\zeta} z'dZ'(z') = 1,$$

where the fact that $E[e'] = 1$ has been used. This implies that the aggregate wage rate, w, is given by

$$w = \lambda \left(\frac{\kappa}{\widetilde{\widetilde{r}}}\right)^{\kappa/(\zeta+\lambda)} \left[\int z'dZ'(z')\right]^{\zeta/(\zeta+\lambda)} = \lambda \left(\frac{\kappa}{\widetilde{\widetilde{r}}}\right)^{\kappa/(\zeta+\lambda)} \mathbf{z}'^{\zeta/(\zeta+\lambda)}. \tag{20}$$

The wage rate, w, depends on the mean of the end-of-period productivity distribution across firms, $\mathbf{z}' \equiv \int z'dZ'(z')$.

Next, suppose that there is free entry by agents into the market for d-type patents. This dictates that the price q will be determined by

$$q = A(\mathbf{z}). \tag{21}$$

To complete the description of a symmetric balanced growth equilibrium, the distribution over propinquity for patent agents, or $D(x)$, must be specified. It is uniform in a symmetric equilibrium. Recall that a firm's permanent location in the technology space is represented by a point on the circle. Think about a buyer located at the top of the circle. Suppose that a set of firms on some tiny arc jk to the left of top are selling patents of mass λ that are of distance between 0 and ε away from the top. Now take any other arc lm of equal length even further to the left of top. The start of this second arc has distance $d(j,l)$ from the start of the first one. In a symmetric equilibrium there will be on the second arc, for all practical purposes, an identical set of firms selling patents of mass λ that are of distance between $d(j,l)$ and $d(j,l) + \varepsilon$ away from the top.

2.5.1 Some Features of a Balanced Growth Path

Along a balanced growth path, consumption, investment, output, profits, wages, and the selling and buying prices for d-type patents will all grow at a constant rate. Also, the interest factor and rental rate on capital are constant. Assuming that this is the case, then it is easy to deduce from (20) that wages must grow at the gross rate $\mathbf{g}^{\zeta/(\zeta+\lambda)}$. Aggregate output and profits will grow at this rate too, as can be inferred from (7). Given the assumption that tastes are isoelastic, the interest factor and rental rate on capital are given in standard fashion by

$$r = \beta/\mathbf{g}^{\varepsilon\zeta/(\zeta+\lambda)}, \tag{22}$$

and

$$\widetilde{r} = \mathbf{g}^{\varepsilon\zeta/(\zeta+\lambda)}/\beta - 1 + \delta, \tag{23}$$

where again ε is the coefficient of relative risk aversion. By substituting the solution for wages, as given by (20), into the demand for labor, (19), it can be seen that a firm's employment is proportional to $z'/\mathbf{z}'$. Since on average one would expect that z' will be growing at the same rate as $\mathbf{z}'$, this suggests that a stationary firm-size distribution exists.

It turns out that along a balanced growth path the indicator functions $I_k(z, x; \mathbf{z})$ and $I_a(z, x; \mathbf{z})$ can represented by simple threshold rules for x that do not depend on either z or $\mathbf{z}$. In particular,

$$I_k(z, x; \mathbf{z}) = \begin{cases} 1 \text{ (keep)}, & x > x_k, \\ 0 \text{ (sell)}, & \text{otherwise}, \end{cases} \quad \text{and } I_a(z, x; \mathbf{z}) = \begin{cases} 1 \text{ (sale)}, & x > x_a, \\ 0 \text{ (no sale)}, & \text{otherwise.} \end{cases}$$

That is, an innovating firm keeps its d-type idea when $x > x_k$ and sells otherwise. Analogously, a sale between a buyer and a patent agents occurs if and only if $x > x_a$.

3 Empirical Analysis

3.1 Data Sources

This section details data sources and variable constructions. For further information, please see Empirical Appendix 9.

NBER-USPTO Utility Patents Grant Data (PDP). The core of the empirical analysis draws from the NBER-USPTO Patent Grant Database (PDP). Patents are exclusionary rights, granted by national patent offices, to protect a patent holder for a certain amount of time, conditional on sharing the details of the invention. The PDP data contains detailed information on 3,210,361 utility patents granted by the U.S. Patent and Trademark Office between the years 1976 and 2006. A patent has to cite another patent when the former has content related to the latter. When patent A cites patent B, this particular citation becomes both a *backward* citation made by A to B and a *forward* citation received by B from A. Moreover, the PDP contains an International Patent Classification (IPC) code for each patent that helps identify where it lies in the technology space.[8] Extensive use of the forward and backward citations are made, as well as the IPC codes assigned to each patent, to determine a patent's location in the technology space, its distance to a firm's location in the technology spectrum, and also to proxy for a patent's quality. The exact methodology followed to construct these measures is detailed below.

[8]The USPTO originally assigns each patent to a particular U.S. Patent Classification (USPC), which is a system used by the USPTO to organize all patents according to their common technological relevances. The PDP also assigns an IPC code to each patent using the original USPC and a USPC-IPC concordance based on the International Patent Classification Eighth Edition.

Patent Reassignment Data (PRD). The second source of data comes from the recently-released USPTO patent assignment files retrieved from Google Patents Beta. This dataset provides detailed information on the changes in patent ownership for the years 1980 to 2011. The records include 966,427 patent reassignments not only due to *sales*, but also due to *mergers, license grants, splits, mortgages, collaterals, conversions, internal transfers*, etc. Reassignment records are classified according to a search algorithm that looks for keywords, such as "assignment", "purchase", "sale", and "merger", and assigns them to their respective categories. Through this process, 99% of the transaction records are classified into their respective groups–see Empirical Appendix 9 for more information.

Compustat North American Fundamentals (Annual). In order to assess the impact of patents and their technological distance on firm moments, such as stock market valuation, the PDP patent data is linked to Compustat firms. The focus is on the balance sheets of Compustat firms between the years 1974-2006, retrieved from Wharton Research Data Services. The Compustat database and the NBER PDP database are connected using the matching procedure provided in the PDP data.

Lex Machina Database on Patent Litigations. The information on litigated patents is obtained from Lex Machina. It is the most comprehensive database on patent litigations since 2000. Lex Machina obtains its data on a daily basis from (i) the administrative database of the United States federal courts, (ii) all United States District Courts' websites, (iii) the International Trade Commission's (EDIS) website, and (iv) the USPTO's websites.[9]

Derwent Litalert Database on Patent Litigations. For litigation information before 2000, the Derwent Litalert Database is used. Further description about this dataset can be found in Galasso, Schankerman and Serrano (2013).

Carnegie-Mellon Survey (CMS) on Industrial R&D. The sector-level licensing information is drawn from CMS. This dataset is one of the rare R&D surveys in the U.S. that contains information on the licensing activities of firms. CMS contains 1,478 randomly selected R&D labs of manufacturing firms, stratified by three-digit SIC industry codes. All labs are located in the U.S. In the survey the firms are asked to report the most important reason for applying for their product patent, where one of the answers is "to obtain revenue through licensing." The percentage of firms picking this answer is aggregated to the two-digit SIC industry classification, which results in a sector-level licensing intensity measure. More information can be found in Cohen, Nelson and Walsh (2000).

The empirical analysis requires the construction of a notion of distance in the technology space. For that purpose, the citation patterns across IPC technology fields are utilized. The PDP contains the full list of citations with the identity of citing and cited patents. Since the data also contains the IPC code of each patent, the percentage of outgoing citations from one technology class to another

[9]Further details can be found at: https://lexmachina.com/.

are observable. Using this information, the metric (24) is employed to gauge the distance between a new patent and a firm's location in the technology spectrum.

In what follows, for each empirical fact the best and largest possible sample is used. For instance, for the firm value regressions all patents that are matched to the Compustat sample are utilized. Similarly, to describe the change from seller to buyer, all patents for which the buyer and seller could be uniquely identified are used. Therefore, even though the samples will be varying across different empirical facts, this approach delivers the most reliable results.

3.2 Technological Propinquity

The notion of technological propinquity between a patent and a firm is now formalized. Think about a patent as lying within some technological class. Call this technology class X. Empirically this can be represented by the first two digits of its International Patent Classification (IPC) code. Now, one can measure how close two patents classes, X and Y, are to each other. To do this, let $\#(X \cap Y)$ denote the number of all patents that cite patents from technology classes X and Y simultaneously. Let $\#(X \cup Y)$ denote the number of all patents that cite either technology class X and/or Y. Then, the following symmetric distance metric can be constructed:

$$d(X,Y) \equiv 1 - \frac{\#(X \cap Y)}{\#(X \cup Y)},$$

with $0 \leq d(X,Y) \leq 1$. This distance metric is intuitive. If each patent that cites X also cites Y, this metric delivers a distance of $d(X,Y) = 0$. [Also note that $d(X,X) = 0$.] If there is no patent that cites both classes, then the distance becomes $d(X,Y) = 1$. The distance between two technology classes increases, as the fraction of patents that cite both decreases. Given this metric between technology classes, a distance measure between a patent and a firm can now be constructed.

In order to measure how close a patent is to a firm in the technology spectrum, a metric needs to be devised. For this purpose, a firm's past patent portfolio is used to identify the firm's existing location in the technology space. In particular, the distance of a particular patent p to a firm f is computed by calculating the average distance of p to each patent in firm f's patent portfolio as follows:[10]

$$d_\iota(p,f) \equiv \left[\frac{1}{\|\mathcal{P}_f\|} \sum_{p' \in \mathcal{P}_f} d\left(X_p, Y_{p'}\right)^\iota \right]^{1/\iota}, \tag{24}$$

with $0 < \iota \leq 1$, and where $0 \leq d_\iota(p,f) \leq 1$. In this expression, $\mathcal{P}_f$ denotes the set of all patents that were ever invented by firm f prior to patent p, $\|\mathcal{P}_f\|$ stands for its cardinality, and $d\left(X_p, Y_{p'}\right)$ measures the distance between the technology classes of patents p and p'. Note that $d\left(X_p, Y_{p'}\right) = 0$ when the firm has another patent, p', in the same class as p. Therefore, this metric is defined only

[10] The firm's patent portfolio is defined as all inventions by the firm up to that point in time.

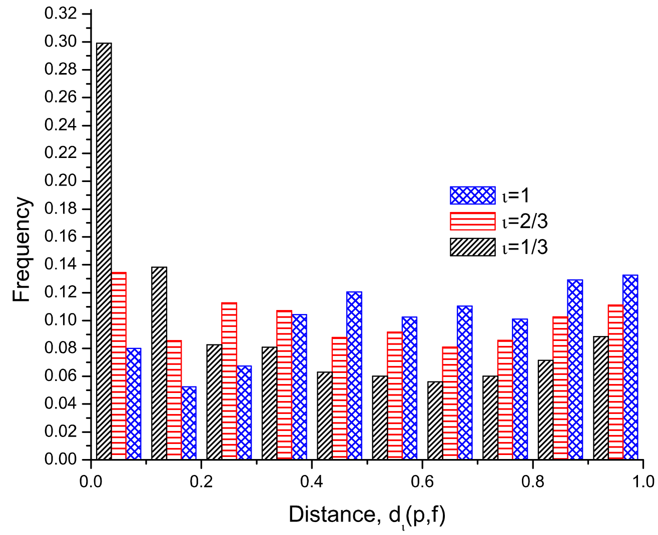


Figure 3: Empirical distance distributions. The figure plots empirical density functions for the distance, $d_\iota(p, f)$, between a patent, p, and a firm's patent portfolio, f, for three values of ι.

for $\iota > 0$. Finally, when $\iota = 1$ the above metric returns the average distance of p to each patent in firm f's patent portfolio: $d_1(p, f) \equiv \|\mathcal{P}_f\|^{-1} \sum_{p' \in \mathcal{P}_f} d\left(X_p, Y_{p'}\right)$, with $0 \leq d_1(p, f) \leq 1$.

The empirical distribution for this notion of distance is displayed in Figure 3 for three values of ι. As can be seen, patents have heterogeneous technological distances to the inventing firms. The intermediate value, $\iota = 2/3$, is chosen for the subsequent analysis.[11]

3.3 Stylized Facts

Next, the empirical findings highlighted in the introduction of the paper are presented. Table 1 provides the summary statistics. Panel A shows the summary statistics of the variables computed using Compustat firms. The distance-adjusted patent stock is constructed in a way such that each patent's contribution to the portfolio is multiplied by its distance to the firm prior to the aggregation. Specifically,

$$\sum_{p \in \mathcal{P}_f} d_\iota(p, f) \times \textsc{quality}(p)$$

where $d_\iota(p, f)$ and $\textsc{quality}(p)$ are the distance and quality terms for patent p. The quality of a patent is measured by the citations it has received from other patents, corrected for truncation and technology class biases using the weights offered in Hall, Jaffe and Trajtenberg (2001).

Panel B reports the summary statistics of the USPTO/NBER patent data. As seen, the average distance between a new patent and its firm is 0.48. The so-called "garage inventors" and firms that

[11] The value chosen for ι does not appear to make much of a difference for the analysis. For example, both the empirical and model simulation results in the paper are more or less the same when either $\iota = 1/3$ or $\iota = 1$.

Summary Statistics

	Observation	*Mean*	*St. Dev*
Panel A. Compustat Facts			
log market value	37,331	5.58	2.30
log employment	39,431	0.75	2.28
log patent stock	41,515	5.68	2.25
log distance-adjusted patent stock	42,269	3.38	4.23
Panel B. USPTO/NBER Patent Facts			
patent quality	2,789,701	12.1	20.7
patent distance	2,565,424	0.48	0.30
litigation probability	2,790,905	0.01	0.10
Panel C. Patent Reassignment Facts			
fraction of patents sold (at least once)	3,210,361	0.16	0.36
number of times a patent is sold	3,210,361	0.19	0.52
conditional duration of patent sale, yrs	421,936	5.48	4.58
litigation and sale probability	2,790,905	0.003	0.06
Panel D. Cumulative Density			
	0 times	*1 time*	*2 times*
number of times a patent is sold	85%	97%	99%

Table 1: Patent quality is measured by the number of patent citations corrected for truncation using the "HJT correction term" from Hall, Jaffe and Trajtenberg (2001). "Portfolio size" is defined as the number of patents that the innovating firm has ever produced by the time of the current innovation. "Transfer duration" is measured by the grant date, with negative durations being dropped.

do not have any existing patents in their portfolio are dropped when patent distance is computed. Panel C lists the summary statistics using patent reassignment data. On average, 15.6% of patents in the sample were traded at least once. The mean time to sell a patent after its grant date is 5.5 years. The average number of trades per patent is 0.2. Panel D shows that 97% of patents are traded at most one time and this number goes up to 99% when the fraction of patents that are traded at most two times are considered. Only a paltry 1.0% of patents involve litigation. The following fact summarizes this section.

Fact 1 *About 15% of patents are sold and it takes about 5.5 years to sell them on average.*

3.3.1 Firm Market Value and Patent-Firm Distance

Are patent-firm distances important when it comes to the relationship between a firm's patent portfolio and its value? In order to answer this question, Table 2 regresses "log market value" in year t on a firm's patent portfolio, its distance-adjusted patent portfolio, and the firm's size in the same year. The regressions also include year and firm fixed effects to rule out firm-specific properties and time trends.

As expected, column 1 shows that the patent portfolio of a firm is positively related to its stock

Firm Market-Value Regressions

	Dependent Variable: log market value					
	(1)	(2)	(3)	(4)	(5)	(6)
log patent stock	0.039***	0.039***	0.040***	0.038***	0.038***	0.069***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)
log dist-adj pat stock	-0.020***	-0.020***	-0.021***	-0.021***	-0.020***	-0.032***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
log employment	0.728***	0.729***	0.735***	0.763***	0.763***	0.692***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)
firm litigation intensity	no	yes	no	no	yes	no
sector litigation intensity	no	no	yes	no	yes	no
sector licensing intensity	no	no	no	yes	yes	no
only renewed patents	no	no	no	no	no	yes
Obs	36,094	36,091	33,094	33,062	33,060	27,526
R^2	0.92	0.92	0.92	0.92	0.92	0.92

Table 2: Compustat, firm-level regressions. Standard errors are in parentheses and *** denotes significance at the 1% level. Firm and year fixed effects and intercept terms are included in all specifications. The last column repeats the regression in the first column but excludes all patents that have not been renewed at least once.

market valuation. Presumably this is because patents are protecting knowledge that is valuable for the firm. More interestingly, a firm's patent portfolio, once adjusted by patent distances, is negatively related to the firm's market value. The coefficient for the distance-adjusted patent stock quantifies the loss of correlation between the patent portfolio and firm value due to the technological mismatch between the firm and its patents. In short, while the non-distance component of the patent portfolio contributes positively, the distance-related component contributes negatively to firm value. In order to interpret the results correctly, consider the ratio of the (negative) coefficient of the distance-adjusted patent stock to that of the unadjusted patent stock. The ratio of the two elasticities is 51.3%. This reflects the relative importance on market value of a shift in the distance-adjusted patent portfolio versus a change in the non-adjusted one. This ratio will be targeted in the simulation. It provides information on the importance of d-type patents relative to n-type ones.

Two factors that have been receiving some attention in the literature recently are licensing and litigation. They could influence a firm's incentives to do R&D, the value of a firm's patent stock, or a firm's decision to buy, keep or sell patents. Licensing is an alternative vehicle for technology transfer. Additionally, litigation might affect a firm's decision to acquire, retain or sell patents. Therefore, controls are introduced for litigation and licensing: Columns 2-4 introduce the fraction of a firm's portfolio that is ever litigated, sector-level litigation intensity (defined as the fraction of litigated patents over total patents in that sector), and sector-level licensing intensity, respectively. Column 5 introduces all these controls at once. All of these alternative specifications show that the benchmark estimates in column 1 are remarkably robust. Last, some patents have little value. To control for this, the last column only includes those patents that were renewed at least once.[12] (Patents must be renewed, at a small fee, in their 3rd, 7th and 11th years.) As can be seen, a lot of patents aren't renewed and purging these patents increases somewhat the impact of the patent stock and distance-adjusted patent stock on the firm's market value. The story remains more or less the same, though, with the relative value of the first two regression coefficients more or less staying fixed. The gist of this section is summarized as follows:

Fact 2 *A patent contributes more to a firm's stock market value if it is closer to the firm in terms of technological distance.*

3.3.2 Patent Sale Decision and Patent-Firm Distance

Does the technological distance of a patent to the firm influence the decision to keep or sell it? In order to conduct this analysis, the indicator variable for whether a patent is sold or kept (=1 if a patent is sold, =0 if not) is regressed on a number of potentially related regressors, including the patent's distance to the initial owner. Table 3 reports the OLS regression results.

Using the full sample, column of Table 3 indicates that a patent is more likely to be sold if it is more distant to the firm. The regression includes controls for size of the patent portfolio of the firm, patent quality, year and firm fixed effects. The coefficient on the distance variable is statistically significant and positive. Considering the average number of patents sold ($\simeq$ 15%) in the time period, the coefficient suggests that a perfectly mismatched patent is 14.1% ($\simeq$ 0.0212/0.15) more likely to be sold to another firm, rather than being kept. Recall also that the definition employed for a sale is quite conservative, in the sense that patent transfers due to mergers and acquisitions are not considered sales, even though the primary motive for these events might be the acquisition of patents. The results are in line with the intuition that a firm is more likely to sell patents that are not a good fit, rather than keeping them, due to the potential gains from trading the patent to a firm that might be better suited to exploit the embedded ideas commercially.

Column 2 controls for the litigation intensity of the technology class, while column 3 controls for the lifetime litigation status of the patent, and column 4 for both simultaneously. The association of distance to a patent sale is unaffected by the presence of these additional controls. Column 5 redoes the first regression but purges those patents that aren't renewed at least once. The effect of distance

[12] Information on patent renewals is obtained from USPTO's U.S. Patent Grant Maintenance Fee Events–for details, see https://eipweb.uspto.gov/MaintFeeEvents/ .

Patent Sale Decision
(Full Sample with Litigation Intensity)

	Dependent Variable (=1, if sold, 0 otherwise)				
	(1)	(2)	(3)	(4)	(5)
distance	2.120***	2.127***	2.126***	2.136***	2.562***
	(0.078)	(0.078)	(0.078)	(0.078)	(0.092)
tech-class litigation intensity	no	yes	no	yes	no
patent litigation dummy	no	no	yes	yes	no
renewed patents only	no	no	no	no	yes
Obs	2,564,305	2,564,305	2,564,305	2,564,305	1,990,998
R^2	0.42	0.42	0.42	0.42	0.44

Table 3: The dependent variable is multiplied by 100 for clarity. All specifications control for the number of citations that each patent received (using the HJT correction term), for the size of the firm's patent portfolio and also for firm fixed effects. Year and intercept terms are included. The coefficients on the non distance terms are not reported to save space and are available upon request. The last column repeats the regression in the first column of the panel but excludes all patents that have not been renewed once. Standard errors are in parentheses and *** denotes significance at the 1% level.

is only slightly more pronounced. The results are also robust to inclusion of the licensing intensity of a sector–see Section 9.6 in the Empirical Appendix for the results. The stark result is that adding additional controls or restricting the sample do not weaken the link between distance and patent sale; to the contrary, they make it more pronounced. The implications of litigation and licensing on the market for ideas will be explored in more detail in Section 6.

Fact 3 *A patent is more likely to be sold the more distant it is to a firm.*

3.3.3 Patent-Firm Distance Reduction Conditional on a Patent Sale

The primary motivation behind considering patent distance as a likely determinant of patent sale decisions is the potential gains from trade that arise if the patent can be sold to a firm that can use it better, which in expectation yields more profits. If this intuition is correct, the distance between the owner firm and the patent is expected to decrease after a patent is sold. Let $d(p, f_b)$ denote the distance of the patent to the buyer firm, and $d(p, f_s)$ to the seller firm. Next, the change in distance, $d(p, f_b) - d(p, f_s)$, is computed. This difference is -0.152 in 1980, the beginning of the sample, with a standard error of 0.049. What this shows is that conditional on a patent sale, the distance between a patent and its owner is significantly decreased. In other words, the mismatch between the idea and the firm owning it is reduced. The effect is economically large. Considering that the average measure for distance is 0.481, the average reduction in distance is approximately 32% ($\simeq 0.152/0.481$) of the average distance. The average distance reduction in the whole sample is 16% and this number goes up to as high as 49% in 2006, which is the end year of the sample.

Fact 4 *A patent is technologically closer to the buying firm than to the selling firm.*

3.4 Tacking on a Market for n-type Patents

To append a market for n-type ideas onto the model, recall that a firm obtains an n-type idea with probability $\mathfrak{p}$. This can arise in one of two ways: either the firm develops an n-type idea or it purchases one. Let a firm that develops an n-type idea sell it with probability $\mathfrak{p}_s$. Likewise, assume that a firm that fails to come up with an n-type idea will purchase one with probability $\mathfrak{p}_b$. Suppose that the market for n-type ideas clears instantaneously every period. This implies that $\mathfrak{p}\mathfrak{p}_s = (1-\mathfrak{p})\mathfrak{p}_b$, so that $\mathfrak{p}_b = \mathfrak{p}_s\mathfrak{p}/(1-\mathfrak{p})$. Adding a market for n-type patents onto the above structure does *not* alter the model's solution for a symmetric balanced growth path. This is discussed further in the Theory Appendix, Section 8.2.

In the U.S. data the distance between a patent and its owner's line of business shrinks on average upon sale; i.e., a patent is closer to the buyer than the seller. This is not true empirically for all patent sales. The presence of n-type patents helps the model better capture Fact 4. It is easy to deduce that on average the distance between a d-type patent and its owner would contract in the model by $[1/(1-x_a)]\int_{x_a}^{1} x dx - [1/X(x_k)]\int_0^{x_k} x dX(x)$, since a non-innovating business buys if $x > x_a$ and an innovating firm sells when $x < x_k$. The average distance between an n-type patent and its owner would contract in the model by $\int_0^1 x dx - \int_0^1 x dX(x)$.[13] This is smaller than the number for d-type patents, because $[1/(1-x_a)]\int_{x_a}^{1} x dx > \int_0^1 x dx$ and $[1/X(x_k)]\int_0^{x_k} x dX(x) < \int_0^1 x dX(x)$. Thus, the presence of a market for n-type patents operates to reduce the average shrinkage in distance upon sale between a patent and its owner.

4 Calibration

In order to simulate the model values must be assigned to the various parameters. There are sixteen parameters to pick: β, ε, κ, λ, δ, σ, γ_d, χ, ρ, μ, η, ω, γ_n, $\mathfrak{p}$, $\mathfrak{p}_s$, and $\text{STD}(e')$. A distribution for $X(x)$ needs to be provided as well. As is standard in macroeconomics, some of the parameter values are chosen on the basis of a priori information, while others are determined internally using a minimum distance estimation routine. By selecting some parameters using a priori information the size of the calibration/estimation procedure is reduced. This is important because undertaking

[13] The distance between an n-type patent and its owner has no real effect; i.e., the technology class for an n-type patent is just a label.

calibration/estimation is problematic when there is a large number of parameter values. For the most part, there is either a consensus about what the appropriate values for these parameters are, or the U.S. data speaks directly to them. The selection of parameter values on the basis of a priori information is now discussed.

4.1 The Use of A Priori Information

1. *Capital's and labor's shares of income, κ and λ.* In line with Corrado, Hulton and Sichel (2009) estimates from the U.S. National Income and Product Accounts, capital's and labor's shares of incomes, κ and λ, are set to 25 and 60%. This implies that the profit parameter, as represented by ζ, accounts for the remaining 15%. This is a fairly typical value used in the macroeconomics literature, as is discussed in Guner, Ventura, and Xu (2008).

2. *Depreciation rate for capital*, δ. The depreciation rate of capital is chosen to be 6.9%. This is consistent with the U.S. National Income and Product Accounts.

3. *Survival rate for a patent*, σ. In the U.S. a patent lasts for 17 years. Hence, $\sigma = 1 - 1/(1 + 17)$.

4. *CRRA parameter*, ε. This parameter is taken to be 2, the midpoint between the various estimates reported in Kaplow (2005). This is a common value used in macroeconomics.

5. *Long-run interest rate.* A reasonable value for the long-run interest rate in the U.S. is 6%–see Cooley and Prescott (1995). Now, the long-run growth rate for the U.S. is 2%. Given the value for the economy's long-run growth rate, $\mathbf{g}^{\zeta/(\zeta+\lambda)} = 0.02$, and the coefficient of relative risk aversion, $\varepsilon = 2$, the discount factor, β, is then uniquely pinned down using the equation $\beta = r\mathbf{g}^{\varepsilon\zeta/(\zeta+\lambda)}$– see (22). This is standard procedure for a growth model.

6. *The Empirical Distribution for the Proximity of Patent to a Firm's Technology Class.* The empirical distance distribution for the U.S. displayed in Figure 3, for $\iota = 2/3$, is used for the analysis. Define a measure of propinquity (or closeness) between a patent p and a firm f by $c_\iota(p, f) \equiv 1 - d_\iota(p, f)$, where $d_\iota(p, f)$ is given by (24). The density associated with $c_\iota(p, f)$ is used for $X(x)$. This amounts to just a simple change in units on the horizontal axis in Figure 3. Assume that x is distributed uniformly within each of the ten bins of the histogram. (There is an additional mass point at one.) One might think that a firm will try to invent ideas that are close to its line of business. The calibration strategy forces the concordance of ideas with the inventor's line of business in the model to be congruent with the U.S. data.

7. *R&D Cost Elasticity*, ρ. In order to estimate the elasticity of the R&D cost function, the cost function in the model is inverted to obtain a production function. Then, a regression is run using Compustat data to determine the parameter value, ρ, where the output of the R&D production function is proxied for by citation-weighted patents.

8. *Bargaining power*, ω. The bargaining powers of buyers and sellers are chosen to be equal. This assumption is often imposed in macroeconomic models using Nash bargaining. Unfortunately, there does not seem to be a good way to identify a value for this parameter, either using a priori information or through the calibration/estimation procedure discussed below. Due to the presence of a Romer (1986) type spillover externality in (2), the Hosios condition will not necessarily lead to an efficient matching equilibrium.

Therefore, values for the parameters β, ε, κ, λ, δ, σ, ρ, and ω are imposed using a priori information in line with (1), (2), (3), (4), (5), (7), and (8) without having to solve the model. The distribution $X(x)$ is constructed in line with point (6).

4.2 Minimum Distance Estimation

Values for the remaining parameters, χ, μ, γ_d, η , γ_n, $\mathfrak{p}$, $\mathfrak{p}_s$, and $\text{STD}(\ln e')$, must be assigned. This is done by minimizing the sum of the squares between some data targets, discussed below, and the model's predictions for these targets. The model is highly non-linear in nature. Computing the solution to the model essentially involves solving a system of nonlinear equations, as is discussed in the Theory Appendix, Section 8.1. Therefore, it is not the case that a particular parameter is identified uniquely by a particular data target. By computing the Jacobian of the system the influence of each parameter on the data targets can be gauged. The presentation below uses this Jacobian and other features of the framework to discuss, in a heuristic fashion, how the parameters are identified. The Jacobian is presented in Section 9.7 of the Empirical Appendix. The data targets are listed in (1) to (7) below. Targets (1) to (4) are discussed now.

1. *Long-run growth in output.* In the U.S. output grew at about 2% per year over the postwar period. Intuitively, one would expect the parameter γ_d, which governs how d-type innovations enter the law of motion for a firm's productivity growth (2), should play an important role in determining this. The same is true for the n-type patent parameters, γ_n and $\mathfrak{p}$. The Jacobian confirms that these parameters have a positive impact on growth–see Appendix 9.7 for more detail. The term for the d-type patents, or γ_d, dominates the others. The parameter governing the cost of R&D, χ, has a negative and smaller effect on growth.

2. *The ratio of R&D expenditure to GDP*. U.S. expenditure on research and development is about 2.91% of GDP. What parameters influence this ratio? Again, the parameter γ_d governing the productivity of d-type patents is very important. It increases this ratio because the payoff from R&D rises with γ_d. Not surprisingly, the R&D cost parameter χ has a bearing here, because it directly governs the cost of innovation, as can be seen from (4). Last, the n-type patent parameters γ_n and $\mathfrak{p}$ are negatively associated with this ratio. They increase GDP growth without the need to do R&D.

3. *Fraction of patents sold*. About 16% of patents are sold in the U.S., as catalogued in Table 1. The parameters governing the matching function, μ and η, govern how easy it is to sell a d-type patent. They are important in determining this ratio. The parameters, $\mathfrak{p}$ and $\mathfrak{p}_s$, governing the arrival and sales rate for n-type patents are also important, although the dependence here is of a mechanical nature.

4. *Duration until a sale*. The entire empirical frequency distribution for the duration of a sale is targeted–see Figure 4.[14] In particular, the calibration procedure tries to minimize the sum of the squared differences between the empirical distribution and the analogue for the model. It takes about 5.34 years on average to sell a patent. The coefficient of variation around this mean is 0.84. So, there is considerable variation in sale duration. The parameters governing the matching function, μ and especially η, are obviously central here. This can be seen from equation (27) in Section (8.1) of the Theory Appendix, which specifies the odds that a patent agent will find a buyer. These parameters also influence the spread in duration.

4.2.1 Indirect Inference

The data targets (5) and (6) discussed below derive from the firm-level panel-data regressions presented in Section 3. As was mentioned, computing the equilibrium solution for the model essentially involves solving a system of nonlinear equations, as the Theory Appendix, Section 8.1, makes clear. Undertaking the indirect inference involves an additional step. Here a Monte Carlo simulation is

[14] Different criteria can be used for dating when an idea is born. One could use the application date instead of the grant date since some patents are sold before they are granted. An alternative would be to use the first time that another inventor builds on this invention (as measured by the first citation that a patent receives). This reflects the time it that took for others to learn about the idea. Last, it is possible that excluding more recent observations might prevent the confounding effects of a potential truncation bias. This occurs because patents toward the end of sample have less time to be sold. Repeating the analysis using these three new sale duration distributions does not change the main findings. Again, see NBER Working Paper 19763 for the details.

undertaken on a panel of 30,000 firms for 30 periods (to replicate the number of periods in the data). This is used to estimate the panel-data regressions analogues for the model that correspond with the ones estimated from the U.S. data, which are presented in Table 2. The last data target (7) is the reduction in distance that occurs between a firm and a patent following a sale, discussed in Section 3.3.3. The Monte Carlo is also used to compute a model analogue for this estimate.

5. *Relative strength of the patent stock versus the distance-adjusted patent stock on a firm's market value.* This is estimated from the micro data–Table 2, column 1. It is measured by the ratio of the coefficient on log distance-adjusted patent stock to log patent stock. This target plays a significant role in identifying the size of the distance related term, γ_d, relative to the non-distance related ones, γ_n and $\mathfrak{p}$, in the law of motion for productivity (2). The former has a positive impact on this ratio, while the latter have negative ones. The matching function parameters, μ and η, also have an influence on this target because they affect the value of a d-type patent. Similarly, so does the cost of doing R&D, χ. Last, the probability of selling an n-type patent, conditional upon its arrival, $\mathfrak{p}_s$, affects this statistic. The higher the likelihood that an n-type patents is sold, and therefore that it is not used production, the less impact it will have a firm's market value. This results in d-type patents mattering more for market value relative to n-type ones–again, the detail is in the Jacobian presented in Appendix 9.7.

6. *Relative strength of the patent stock versus employment on a firm's market value.* This, too, is estimated from the micro data–Table 2, column 1. It is measured as the ratio of the coefficient on log patent stock to the coefficient on log employment. The (inverse of this) ratio can be thought of as measuring the impact of an increase in the patent stock on employment, holding fixed the firm's market value. In the model there are two reasons a firm's market value may rise relative to other firms. Its long-run productivity, z', may have increased relative to average long-run productivity, $\mathbf{z}'$, or it may have realized a favorable value for the temporary production shock, e'. This ratio identifies the standard deviation of the firm-specific idiosyncratic production shock, $\text{STD}(\ln e')$. Without the e' shock, employment would be a perfect predictor of relative productivity, $z'/\ \mathbf{z}'$–see (42). Introducing the e' shock breaks this one-to-one correspondence. The parameter $\text{STD}(\ln e')$ has no impact on the other data targets. The parameter γ_n governing the productivity of n-type patents also affects this ratio. As γ_n rises employment becomes a better predictor of a firm's market value, so it impinges on this ratio in a negative way. An increase in γ_d does not work the same way as it results in more d-type ideas, which makes the patent stock a better predictor of market value.

Parameter Values

Parameter Value	*Description*	*Identification*
$\beta = 0.98$	Discount factor	A priori information
$\varepsilon = 2.00$	CRRA parameter	A priori information
$\kappa = 0.25$	Capital's share	A priori information
$\lambda = 0.60$	Labor's share	A priori information
$\delta = 0.07$	Depreciation rate	A priori information
$\sigma = 0.94$	Patent survival rate	A priori information
$\gamma_d = 0.25$	Distance-related productivity	Calibration/Estimation
$\chi = 0.83$	Cost of R&D	Calibration/Estimation
$\rho = 3.00$	R&D cost elasticity	A priori information
$\mu = 0.50$	Matching function, exp	Calibration/Estimation
$\eta = 0.09$	Matching function, const	Calibration/Estimation
$\omega = 0.50$	Bargaining power	Imposed
$X(x)$	Proximity distribution	A priori information
$\gamma_n = 0.18$	Non-distance related productivity	Calibration/Estimation
$\mathfrak{p} = 0.17$	Pr(n-type idea)	Calibration/Estimation
$\mathfrak{p}_s = 0.47$	Pr(sell n-type patent|arrival)	Calibration/Estimation
$\text{STD}(\ln e') = 0.07$	Production shock, std	Calibration/Estimation

Table 4: The parameter values used in the baseline simulation.

7. *Distance reduction upon sale–all patents.* Section 3.3.3 presents an estimate (-0.152) from the the micro data on the average difference between a buyer's and seller's technological propinquity for a patent.[15] This estimate is targeted and helps to discipline the relative importance d- and n-type patents. As is discussed in Section 3.4, the arrival rate of n-type ideas and the probability of selling them, or $\mathfrak{p}$ and $\mathfrak{p}_s$, are central here. They operate to reduce the observed amount of distance reduction since the sale of these patents does not depend upon technological propinquity. This is shown by the Jacobian of the system. Additionally, the parameters of the matching function, μ and η, influence the model's ability to hit this target. More efficient matching implies a larger reduction in distance.

To highlight a central aspect of the calibration procedure, note that a key goal of this research here is to quantify the importance of the patent market for eliminating the misallocation of ideas across producers. Two considerations come into play: the importance of technological propinquity between a patent and a producer (or γ_d) and the efficiency of the market for ideas (or η). A low volume of patent sales could occur either because technological propinquity is not very important (but the patent market is still efficient) or because the patent market is inefficient (but technological propinquity is important). The above micro data is used to identify both of these channels. At the

[15] The quantitative results do not change in a material way when the mean of the averages over all years in the sample is used instead.

Calibration Targets

Target	*U.S. Data*	*Model*
Long-run growth in output	2.00%	2.08%
Ratio of R&D expenditure to GDP	2.91%	1.96%
Fraction of patents that are sold	15.6%	16.6%
Average duration until a sale (fit entire distribution)	5.48 yrs.	6.20 yrs.
Sale duration, c.v. (fit entire distribution)	0.84	0.71
COEF(dist-adj pat stock)/COEF(pat stock)	-0.511	-0.587
COEF(pat stock)/COEF(empl)	0.054	0.054
$d(p, f_s) - d(p, f_b)$, all sold	0.152	0.165

Table 5: In the calibration the full sales duration distribution (17 points) is targeted. The above table just reports the mean and the coefficient of variation for this distribution as summary measures.

risk of sounding repetitive, the firm market-value regressions in Table 2 are used to speak to the size of γ_d. Since firm fixed effects are included in these regressions, there is a strong sense in which changes in the distance-adjusted patent stock are being tied to the firm market value. Therefore, reproducing similar regression results using the model-generated data (in particular the relative size of the log distance-adjusted patent stock to the log patent stock coefficients) helps identify γ_d. Matching up the model's output with the micro data on the fraction of patent sold, average sale duration, and the difference between the buyer's and seller's technological propinquities pins down η. The efficiency of the market for ideas plays a very important in the analysis and is analyzed in detail in Section 5.

It is well known that patents show big differences in terms of their qualities which could also affect their sales. A reasonable belief might be that a small fraction of patents are highly valuable while the median one is not. To take quality heterogeneity into account, all regressions control for patent citations as a proxy for patent quality. So, the empirical analysis attempts to purge concerns about patent quality from the stylized facts.[16]

The upshot of the calibration procedure is displayed in Tables 4 and 5. Figure 4 shows, for both the data and model, the frequency distribution over the duration for a sale. As can be seen, it appears to be harder to affect a sale in data than in the model.

[16] Alternatively, one could introduce quality into the model. In particular, every idea could have a quality component drawn from some distribution. Now, the decisions to buy and sell patents would be a function of distance and quality (in addition to the aggregate state variable). Perhaps the distribution governing quality could be mapped into the empirical distribution for patent citations. Doing this would significantly complicate the analysis, but could be a fruitful avenue for future research.

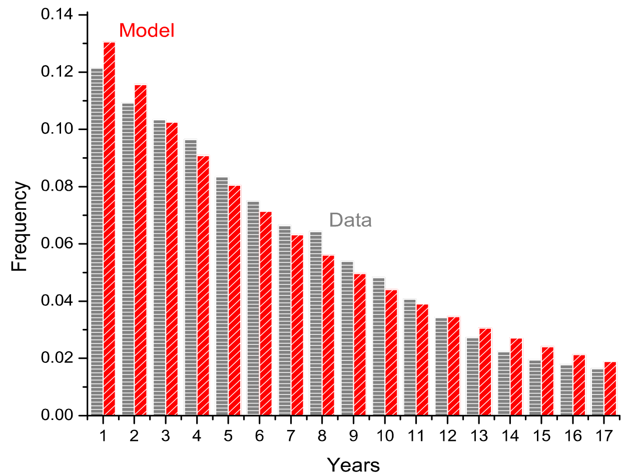


Figure 4: Sales duration distribution, data and model

5 Findings

The importance of a market for patents will be gauged. There are two sources of inefficiencies in the model. The first one is the usual Romer (1986) knowledge externality. Each single innovation raises the aggregate knowledge stock in society, which benefits the future generations that stand on the shoulders of former giants through $\mathbf{z}$ in (2). The second source of inefficiency emerges due to *matching frictions*, which is of particular interest here. To analyze the latter, various experiments that change the efficiency of the market for d-type patents will be entertained. The efficiency of the market for d-type patents is increased in stages. First, the market is shut-down by setting the meeting rate to zero. Then, an experiment is performed where the meeting rate for matches is allowed to rise. While it may be easier for buyers and sellers to meet now, a seller's idea may still not be well suited for the buyer. The next experiment considers a situation where patent agents can find buyers who are perfect matches for the ideas that they are selling. So, there is no mismatch between buyers and sellers on the patent market. Still, innovating firms produce d-type ideas that are not ideally suited for their own businesses and this injects a friction into the analysis. A patent that is not incorporated into an innovator's production process will only have a finite life on the market. Additionally, it may take time to find a buyer. The final experiment focuses on the case where innovating firms produce ideas that are tailored toward their own production activity. Here ideas are perfectly matched with the developer. The change in welfare from moving from one environment to another is calculated. The metric for comparing welfare will be discussed now.

5.1 Welfare Comparisons

Consider two economies, namely A and B, moving along their balanced growth paths. Aggregate consumption, the gross growth rate, and aggregate productivity for economy A are represented by $\mathbf{c}^A$, $\mathbf{g}^A$, and $\mathbf{z}^A$. Similar notation is used for country B. To render things comparable, start each country off from the same initial position where $\mathbf{z}^A = \mathbf{z}^B = 1$. Now, the levels of welfare for economies A and B are given by

$$W^A = \sum_{t=1}^{\infty} \beta^{t-1} \frac{(\mathbf{c}_t^A)^{1-\varepsilon}}{1-\varepsilon} = \frac{(\mathbf{c}_1^A)^{1-\varepsilon}}{(1-\varepsilon)[1-\beta(\mathbf{g}^A)^{1-\varepsilon}]}, \quad \text{and} \quad W^B = \frac{(\mathbf{c}_1^B)^{1-\varepsilon}}{(1-\varepsilon)[1-\beta(\mathbf{g}^B)^{1-\varepsilon}]},$$

where $\mathbf{c}_1^A$ and $\mathbf{c}_1^A$ are the time-1 levels of consumption in economies A and B. How much would initial consumption in economy A have to be raised or lowered to make people have the same welfare level as in economy B? Denote the fractional amount in gross terms by α (which may be less than one). Then, α must solve

$$\frac{(\alpha \mathbf{c}_1^A)^{1-\varepsilon}}{(1-\varepsilon)[1-\beta(\mathbf{g}^A)^{1-\varepsilon}]} = W^B,$$

so that

$$\alpha = (W^B/W^A)^{1/(1-\varepsilon)}.$$

This is welfare measure is used in all experiments.

5.2 Varying the Contact Rate for Matches, η

The patent market mitigates the initial misallocation of ideas. Still, it takes time to sell a patent as the patent agent may not be able to find a buyer. To understand how this friction in matching affects the economy, it is useful to examine the relationship between the scale factor for the matching function, η, and several key variables.[17] Figures 5 and 6 summarize the results.

The market for d-type patents is shut down when $\eta = 0$. When there is no market, the equilibrium growth rate goes down to 2.02% from from its benchmark value of 2.08%. Shutting down the market results in a welfare reduction of 1.18% in consumption equivalent terms, which is quite sizable. As the contact rate, η, rises it becomes easier to find a buyer for a patent, ceteris paribus. This is reflected in a drop in the length of time that it takes to find a buyer, as the right panel of Figure 5 illustrates. The price that an innovating firm receives for a patent, q, rises accordingly–see the left panel of Figure 6. As the price moves up an innovating firm becomes choosier about the patents that it will keep. Figure 6, right panel, illustrates how an innovator's cutoff for selling, x_k, rises with η.

[17] The relationship between $\mathbf{g}$ and η is highlighted in the simplified model developed in Supplemental Appendix 12.

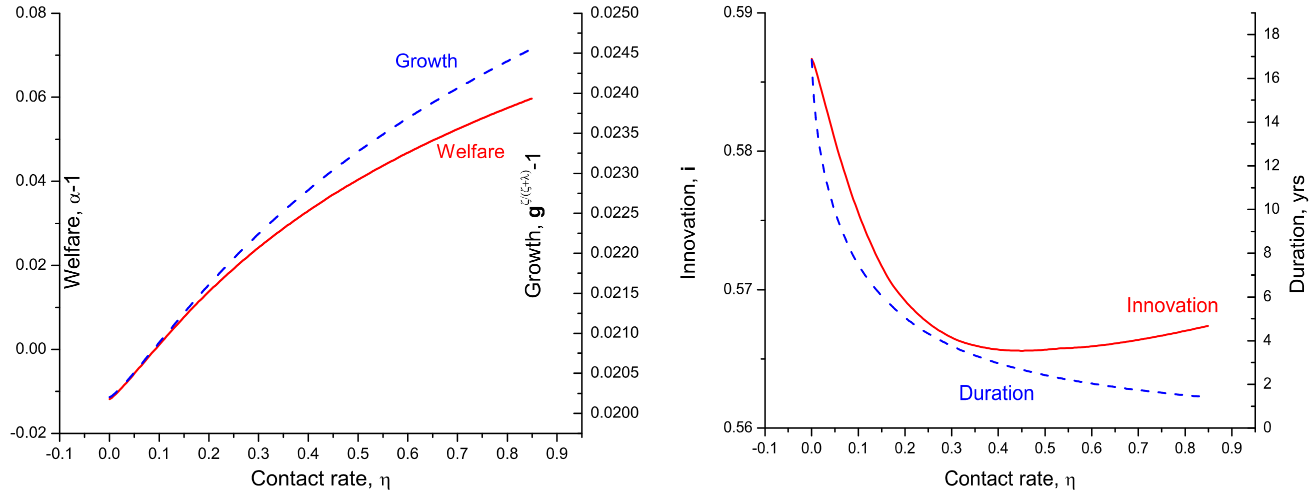


Figure 5: The impact of an increase in the contact rate on duration, innovation, growth and welfare.

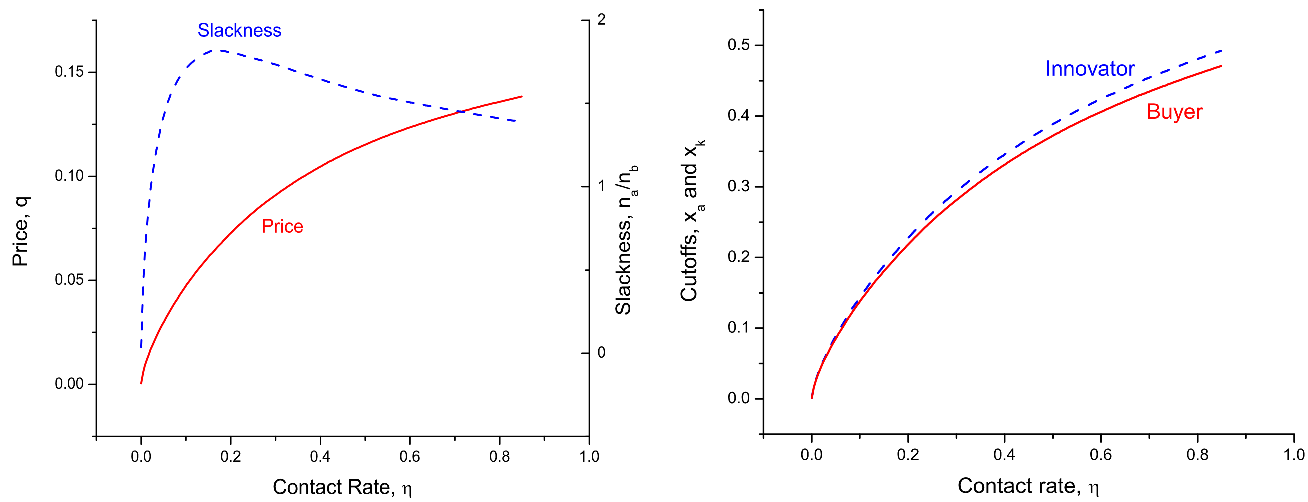


Figure 6: The impact of an increase in the contact rate on the price for the innovator, slackness and the cutoffs.

(Recall that better patents are associated with higher values for the propinquity metric.) Similarly, buyers become pickier about the patents that will they purchase so that x_a moves up with η.

The rate of innovation, $\mathbf{i}$, does not change much. It falls as η starts to rise since the consequences of failing to innovate are now lessened, because it will be easier for a firm to buy a patent. The high price for patents begins to spur innovation at higher levels of η. Market slackness, n_a/n_b, has an interesting $\cap$ shape, which is displayed in the left panel of Figure 6. When $\eta = 0$ the patent market is essentially closed as no innovators will want to sell their ideas. The number of prospective buyers is $1 - \mathbf{i}$. As η starts to rise so does the number of innovators that want to sell their ideas. This increases the flow of new patents into the patent market and results in n_a/n_b moving upwards. As the rate of innovation, $\mathbf{i}$, declines the number of prospective buyers, $1 - \mathbf{i}$, rises. This force operates to reduce n_a/n_b. Additionally, as the contact rates increases the market becomes more efficient. It is easier for a seller to find a buyer, ceteris paribus. This works to reduce the stock of sellers.

Growth increases along with efficiency in matching, despite the reduction in the number of new ideas–see the left panel of Figure 5. So does welfare. If the efficiency of the market was at its extreme (the minimum value for η that results in all buyers meeting a patent agent with probability 1), growth would go up to 2.46% and welfare would be 5.97% higher than the calibrated economy. The upshot is that the market for patents plays an important role in the economy.

5.3 Perfectly Directed Search

A second source of inefficiency in the model is the random search technology used in the d-type patent market. In the baseline model, conditional upon a meeting between a buyer and a patent agent, the propinquity of the idea to the firm is drawn from a uniform distribution. Instead imagine a perfectly directed search structure, where patent agents are able to target the segment of the economy that exactly matches the patent they want to sell. In such a case, whether or not a patent agent meets a buyer is still a probabilistic event governed by the matching function. The propinquity between the patent and the firm would be nonstochastic and equal to unity; in other words, a perfect match. The level of welfare in this alternative economy is 1.94% higher than in the baseline one. The output growth rate increases slightly from 2.08 to 2.19%, despite a small decline in the innovation rate. The fraction of patents sold increases from 16.6 to 19.9%. Last, a decomposition of growth reveals that fraction of growth due to patents sold moves up from 18.9 to 26.6%.[18] Table 6 summarizes the results (where the baseline model is labeled BM and PDS refers to the perfectly directed search structure).

5.3.1 Perfectly Directed Search with a High Contact Rate

Now, redo the above experiment with perfect directly search while also using a high contact rate for matches.[19] The results are reported in Table 6 (under the column labeled PDSwHC). Output growth is now much higher at 3.05%, even though innovation is slightly lower than in the baseline model.

[18] Appendix 8.1 shows that $\mathbf{g} - \mathbf{1} = \gamma_d \mathbf{i} \int_{x_k}^1 x dX(x) + \gamma_d (1-\mathbf{i}) m_b(\frac{n_a}{n_b}) \int_{x_a}^1 x dx \ + \gamma_n \mathfrak{p}$, where $\mathbf{i}$ is the aggregate rate of innovation. Note that there are three terms on the right side. The first term can be used to measure the contribution to growth from the distance-related ideas that firms keep, the second from the ones that they sell. The third term gives the growth arising from non-distance related ideas. This term can be further decomposed as $\gamma_n \mathfrak{p} = \gamma_n [\mathfrak{p}(1-\mathfrak{p}_s) + (1-\mathfrak{p})\mathfrak{p}_b]$, where the first term in brackets gives the contribution from non-distance-related patents kept and the second from the ones sold.

[19] The contact rate, η, is set high enough that all buyers meet a patent agent with probability 1.

Thought Experiments

	BM	*PDS*	*PDSwHC*	*PI*
Output growth rate, %, $(\mathbf{g}^{\zeta/(\zeta+\lambda)} - 1) \times 100$	2.08	2.19	3.05	3.38
Innovation rate, $\mathbf{i}$	0.58	0.56	0.57	0.61
Welfare gain, $\alpha - 1$	0.00	0.02	0.14	0.18
Fraction of d-type patents sold	0.17	0.20	0.68	0
Growth from d-type patents sold	0.19	0.27	0.73	0

Table 6: The first column of results is for the baseline model (BM). Perfectly directly search (PDS) is shown in the second column where a patent sold is a perfect match for the buyer (x=1). In the third column (PDSwHC) there is perfectly directed search, plus there is a high contact rate between patent agents and buyers. All innovating firms draw the perfect idea (x=1) in the last column (PI). The figures in the first row (only) are in percent.

This reflects a reduction in misallocation. As can be seen, now most patents are sold. Economic welfare is 14.3% higher.[20]

Figure 7 gives the upshot from the experiments that have been conducted so far. It shows how the cumulative distribution function for the propinquity of new ideas to firms, or for x, changes across the various experiments. First, firms in the U.S. data produce ideas that are not well suited for their own lines of business, as can be seen from the distribution labeled "Empirical". (Recall that a higher value for $x \in [0, 1]$ indicates that an idea is better suited for the firm's business activity.) In the baseline model, a firm is free to sell such an idea. A firm that fails to innovate can try to buy one from another firm. This leads to a better distribution of ideas, as is reflected in the distribution function for the baseline model after transactions on the market for patents have been consummated. The distribution function for the baseline model stochastically dominates, in the first-order sense, the empirical distribution. When the contact rate for matching is high it is relatively easy to consummate a patent sale. The distribution for x improves–see the histogram labeled "High Contact Rate", which stochastically dominates the one for the baseline model. Of course, if search could be perfectly directed things would be better still–"High Contact w Directed Search", which stochastically dominates all other distributions.

Note that not all firms sell their patents, even though they are not perfectly matched with their ideas. This occurs because are still some frictions left in the patent market. First, there are more sellers than buyers on the market, so not all patents will be immediately sold. Second, patents have a finite life on the market and hence suffer some depreciation. Both these factors imply that the price at which a firm can sell a patent, q, will be less than what it is worth to a perfectly matched

[20] This large welfare gain derives solely from the large increase in growth, $\mathbf{g}^{\zeta/(\zeta+\lambda)}$, given the assumed form of preferences over consumption, as can be gleaned from Section 5.1. That is, if there is a large increase in growth then this form of preferences will always show a large increase in welfare (when $\varepsilon = 2$, which is a standard value).

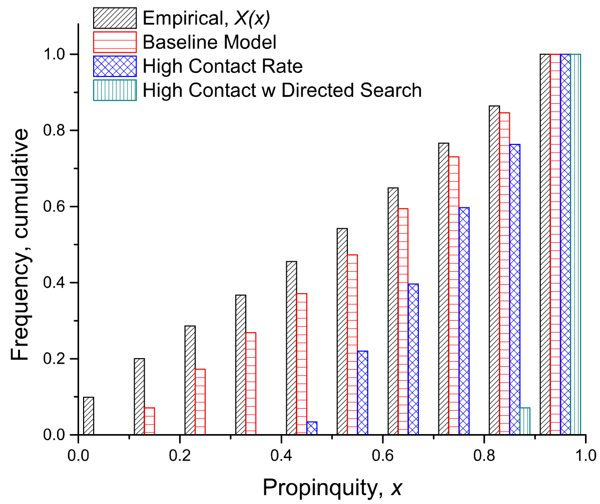


Figure 7: Misallocation of Ideas. The graph plots the cumulative distribution functions for x. A higher value for x, measuring propinquity, implies that an idea is better suited for a firm.

firm.

5.4 Removing the Misallocation of Ideas

The central inefficiency in the framework derives from the fact that firms develop ideas that are imperfect matches for the own production processes. The presence of a market for patents mitigates this problem. Suppose that an innovating firm always comes up with a d-type idea that is a perfect match for its production process. That is, let each innovating firm always draw $x = 1$. This case is summarized in Table 6 (under the column labeled PI). In this situation, the economy could increase its growth rate from 2.08 to 3.38%, a big jump. Welfare would increase by 17.8%. This illustrates that the frictions arising from mismatches in innovation are large.

6 Quantitative Extensions: Licensing and Litigation

When it comes to technology transfer and the market for ideas, two important concerns about patenting and the market for ideas deserve additional attention, namely licensing and litigation. Licensing provides an additional mechanism for transferring ideas. By limiting attention to patent sales, a fear might be that the analysis overstates the amount of misallocation in the market for ideas. A firm may buy or keep a patent to prevent litigation. This does not increase the firm's productivity in a technological sense. Hence, the value of patents on a firm's productivity may be overestimated.

Results with Licensing and Low-Ligation Sectors

Panel A: Calibration Targets				
	Licensing		*Low-Litigation*	
	U.S. Data	Model	U.S. Data	Model
Growth in Output, %	2.00	2.03	2.00	2.10
Ratio of R&D expenditure to GDP, %	2.91	1.81	2.91	1.98
Fraction of Ideas that are sold, %	20.6	20.5	16.4	17.6
Average duration until a sale, yrs.	5.48	6.05	5.94	6.35
Sale duration, c.v.	0.84	0.72	0.78	0.70
COEF(dist-adj pat stock)/COEF(pat stock)	-0.511	-0.605	-0.568	-0.610
COEF(pat stock)/COEF(empl)	0.054	0.058	0.052	0.052
$d(p, f_s) - d(p, f_b)$, all sold	0.152	0.161	0.136	0.143
Panel B: Impact of Shutting Down the Market for Ideas ($\eta = 0$)				
		Benchmark	*Licensing*	*Low-Litigation*
Δ in Growth (percentage pt.)		-0.06	-0.07	-0.06
Δ in Welfare, %		-1.18%	-1.40%	-1.12%

Table 7: Results for both the data and model when ideas can be also transferred via licensing and when the analysis is restricted to low-litigation sectors.

6.1 Licensing

Arora and Ceccagnoli (2006) report that licensing intensity in the U.S. is around 5%. The goal here is to understand the quantitative implications of licensing in the current setting. Zuniga and Guellec (2009) conduct a survey on firms that license out their patents and analyze the obstacles to licensing. The most frequent problem reported by firms was that "identifying (a) partner is difficult." This shows that search frictions, which are highlighted in the model of the patent market developed here, seem to apply to the licensing market as well. Licensing could have many other purposes than pure technology transfer, such as deterring entry. To the extent that licensing is used as a substitute arrangement for patent sale, the previous analysis might have underestimated the liquidity in the market for ideas and generated too much search frictions. In order to take this substitutability into account, assume that all the licensing arrangements are for the purpose of technology transfer. Hence in what follows, assume that the overall turnover in the market for ideas is $20.6\% = 15.6\% + 5\%$. The model is recalibrated and simulated using this number. Table 7 reports the results. The model matches the data well when it is recalibrated to allow for a larger number of ideas to be transferred. Not surprisingly, a shutdown in the market for ideas leads now to a bigger welfare loss (1.40 versus 1.18%). As before, the reduction in growth is still small, but slightly higher (a loss of 0.07 versus 0.06 percentage points). Again, the small loss in growth is due to the fact that the rate of innovation rises when the market for ideas is closed, as was shown earlier in the right panel of Figure 5.

6.2 Litigation

Patent litigation could also lead to patent sales for reasons not necessarily related to technology transfer [Galasso, Schankerman and Serrano (2013)]. To begin with, it is useful to get a sense of the share of patents that are ever litigated in the sample employed here. Using the Derwent and Lex Machina databases, Table 1 shows that about 1.0% of patents involve litigation. Furthermore, when patents that are both ever litigated and ever sold during their lifetime are considered, the share drops down to 0.3%. Hence, among sold patents, only 2% (=0.3/15.6) are ever litigated. Given these small shares, it may seem unlikely that litigated patents could have a major impact on the quantitative results.

As Galasso, Schankerman and Serrano (2013) emphasize, however, the threat of litigation might be very important in the sale decision, even if in practice, few litigations are actually observed. In order to exclude this potential channel, the analysis is redone, focusing exclusively on sectors with very low litigation intensity. All the data targets except the U.S. growth rate and R&D expenditures to GDP are recalculated using patent and firm observations that have a litigation intensity below the mean of the pertinent sample. Indeed, sectors have a lot of heterogeneity in terms of the litigations observed, and a sector's litigation intensity might be a good indicator for the propensity of a given patent to be litigated.

Table 7 provides the new estimates and the welfare gain from the market. Note that the affected data targets change only slightly. These changes occur from restricting the micro data to the low-litigation sectors. The model still fits very well. The welfare loss for shutting down the market for ideas is now a bit smaller (1.12 versus 1.18%). The upshot is that focusing on low-litigation sectors does not affect the analysis in a material way.

7 Conclusions

A model of the market for patents is developed here. Each period a firm conducts research and development. This R&D process may spawn new ideas. Some of the ideas are useful for a firm's line of business, others are not. A firm can patent and sell the ideas that are not. The fact it can sell ideas provides an incentive to engage in R&D. Likewise, firms that fail to innovate can attempt to buy ideas. This allows a firm to grow its business. This reduces the incentive to do R&D. The efficiency of the patent market for matching ideas with firms has implications for growth. These are examined here.

The empirical analysis, drawing on the NBER-USPTO patent grant database and patent reassignment data available from Google Patents Beta, establishes five useful facts. First, somewhere between 15 and 20% of patents are sold. Second, it takes on average 5.48 years to sell a patent. Third, a firm's patent stock contributes more to its market value the closer it is to the firm in terms of average technological distance. Fourth, a patent is more likely to be sold the more distant it is to a firm's line of business. Fifth, when a patent is sold it is closer to the buyer's line of business than to the seller's. The empirical analysis attempts to control for licensing and litigation. These five facts suggest that a market for patents may play an important role in correcting the misallocation of ideas across firms. It may also influence a firm's R&D decision.

The developed model is calibrated to match several stylized facts characterizing the U.S. data, such as the postwar rate of growth, the ratio of R&D spending to GDP, the fraction of patents sold, and the empirical sale duration distribution. Additionally, some micro-level facts from panel data regressions are targeted using an indirect inference strategy. Specifically, the importance of distance in a firm's patent portfolio for determining the firm's market value and the reduction in distance between a patent and its owner upon a sale are zeroed in on. The importance of a market for selling patents is then assessed. This is done by conducting a series of thought experiments where the market is first shut down and then the efficiency of the patent market is increased successively. The efficiency of this market is important for economic growth and welfare. Last, economic growth and welfare increase along with the bargaining power of the seller. As the seller's bargaining power rises more of the return from a new idea is assigned to the innovator. Given the spillovers in the growth process that are present in the model, this is beneficial.

The new NBER patent reassignment data opens new and exciting directions for future research on innovation and technological progress. One direction is the analysis of optimal patent policy that not only considers the monopoly distortions and innovation incentives, but also takes into account the possibility of trading ideas through patents. Another direction is the analysis of firm dynamics when patents are not only produced in-house, but also purchased from others. Finally, the role of financial frictions is also a new and important channel that could impact the (mis)allocation of ideas. These are all very exciting and important aspects of technological progress that await further research.

References

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

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

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

Arora, A., and M. Ceccagnoli (2006): "Patent Protection, Complementary Assets, and Firms' Incentives for Technology Licensing," *Management Science*, 52, 293-308.

Benhabib, J., J. Perla, and C. Tonetti (2014): "Catch-up and Fall-back through Innovation and Imitation," *Journal of Economic Growth*, 19, 1–35.

Chiu, J., C. Meh, and R. Wright (2011): "Innovation and Growth with Financial, and Other, Frictions," Working Paper 17512, NBER.

Cohen, W.M., R.R. Nelson, and J.P. Walsh (2000): "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," Working Paper 7552, NBER.

Cooley, T.F., and E.C. Prescott (1995): "Economic Growth and Business Cycles," in *Frontiers of Business Cycle Research*, ed. by T.F. Cooley. Princeton, NJ: Princeton University Press, 1–38.

Corrado C., C. Hulten, and D. Sichel (2009): "Intangible Capital and U.S. Economic Growth," *Review of Income and Wealth*, 55, 661-685.

Figueroa, N., and C.J. Serrano (2013): "Patent Trading Flows of Small and Large Firms," Working Paper 18982, NBER.

Galasso, A., M. Schankerman, and C.J. Serrano (2013): "Trading and Enforcing Patent Rights," *RAND Journal of Economics*, 44, 275-312.

Gort, M. (1962): *Diversification and Integration in American Industry*. Princeton, NJ: Princeton University Press.

Guner, N., G. Ventura, and Y. Xu (2008): "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 11, 721-744.

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

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

Hochberg, Y.V., C.J. Serrano, and R.H. Ziedonis (2014): "Patent Collateral, Investor Commitment, and the Market for Venture Lending," Working Paper 20587, NBER.

Hsieh, C.-T., and P.J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403-1448.

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

Kaplow, L. (2005): "The Value of a Statistical Life and the Coefficient of Relative Risk Aversion," *Journal of Risk and Uncertainty*, 31, 23-34.

Kerr, W.R., and S. Fu (2006). "The RAD-Patent-LRD Mapping Project," Technical Paper, U.S. Bureau of the Census.

König, M.D., J. Lorenz, and F. Zilibotti (2012): "Innovation vs. Imitation and the Evolution of Productivity Distributions," Discussion Paper 8843, CEPR.

Lamoreaux, N.R., and K.L. Sokoloff (2003): "Intermediaries in the U.S. Market for Technologies, 1870-1920," in *Finance, Intermediaries, and Economic Development*, ed. by S. Engerman et al. Cambridge, U.K.: Cambridge University Press, 209-246.

Lucas, R.E. Jr., and B. Moll (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122, 1-51.

Perla, J., and C. Tonetti (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122, 52-76.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11, 707-720.

Romer, P.M. (1986): "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94, 1002-1037.

Serrano, C.J. (2010): "The Dynamics of the Transfer and Renewal of Patents," *RAND Journal of Economics*, 41, 686-708.

Serrano, C.J. (2015): "Estimating the Gains from Trade in the Market for Patent Innovation: Evidence from the Transfer of Patents," Unpublished Manuscript, Universitat Pompeu Fabra.

Zuniga, M.P., and D. Guellec (2009): "Who Licenses Out Patents and Why? Lessons from a Business Survey," Working Paper 2009/5, Directorate for Science, Technology and Industry, OECD.

8 Theory Appendix

8.1 Balanced Growth

The analysis is restricted to studying symmetric balanced growth paths. The solution to the economy along a balanced growth path will now be characterized.[21] Suppose that mean level of productivity for firms, $\mathbf{z}$, grows at the constant gross rate $\mathbf{g}$. Specify the variables z and $\mathbf{z}$ in transformed form so that $\widetilde{\mathbf{z}} = \mathbf{z}^{\zeta/(\zeta+\lambda)}$ and $\widetilde{z} = z/\mathbf{z}^{\lambda/(\zeta+\lambda)}$. Thus, $\widetilde{\mathbf{z}}$ grows at rate $\mathbf{g}^{\zeta/(\zeta+\lambda)}$ and, on average, so will $\widetilde{z}$. It turns out that $\widetilde{\mathbf{z}}$ (or equivalently $\mathbf{z}$) is sufficient to characterize the aggregate state of the economy along a balanced growth path. It also turns out that the form of the distribution for d-type patent buyers, or G, does not matter.

Proposition 1 *(Balanced Growth) There exists a symmetric balanced growth path of the following form:*

[21] A simplified version of the model with a closed-form solution is presented in Supplemental Appendix 12.

1. *The interest factor and rental rate on capital are given by (22) and (23).*

2. *The value functions for buying, keeping and selling firms have linear forms in the state variables* $\widetilde{z}$ *and* $\widetilde{\mathbf{z}}$. *Specifically,* $B(z;\mathbf{z}) = \mathfrak{b}_1\widetilde{z}+\mathfrak{b}_2\widetilde{\mathbf{z}}$, $K(z+\gamma_d x\mathbf{z};\mathbf{z}) = \mathfrak{k}_1\widetilde{z}+\mathfrak{k}_2(x)\widetilde{\mathbf{z}}$, *and* $S(z;\mathbf{z}) = \mathfrak{s}_1\widetilde{z}+\mathfrak{s}_2\widetilde{\mathbf{z}}$.

3. *The indicator function for an innovator specifies a threshold rule such that* $I_k(z,x;\mathbf{z}) = 1$, *whenever* $x > x_k$, *and is zero otherwise. I.e., an innovating firm keeps its d-type idea when* $x > x_k$ *and sells otherwise.*

4. *The indicator function for a sale between a buyer and the patent agent for a d-type idea specifies a threshold rule such that* $I_a(z,x;\mathbf{z}) = 1$, *whenever* $x > x_a$, *and is zero otherwise. I.e., a sale between a buyer and a patent agents occurs if and only if* $x > x_a$.

5. *The value function for a patent agent has the linear form* $A(\mathbf{z}) = \mathfrak{a}\widetilde{\mathbf{z}}$.

6. *The beginning-of-period value function for a firm has the linear form* $V(z;\mathbf{z}) = \mathfrak{v}_1\widetilde{z}+\mathfrak{v}_2\widetilde{\mathbf{z}}$. *The constant rate of innovation for a d-type idea by a firm is*

$$i = \mathbf{i} = \left\{\frac{1}{\chi}\left[X(x_k)\mathfrak{s}_2 + \int_{x_k}^{1}\mathfrak{k}_2(x)dX(x) - \mathfrak{b}_2\right]\right\}^{1/\rho}. \tag{25}$$

7. *The constant net rate of growth for aggregate productivity is implicitly given by*

$$\mathbf{g} - \mathbf{1} = \gamma_d\left[\mathbf{i}\int_{x_k}^{1} x dX(x) + (1-\mathbf{i})m_b(\frac{n_a}{n_b})\int_{x_a}^{1} x dx\right] + \gamma_n\mathfrak{p}, \tag{26}$$

with the aggregate law of motion (3) taking the simple form

$$\mathbf{z}' = \mathbf{g}\mathbf{z}.$$

8. *The prices for selling and buying d-type patents are*

$$q = \mathfrak{a}\widetilde{\mathbf{z}},$$

and

$$P(z,x;\mathbf{z}) = \left[(1-\omega)\sigma r\mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a} + \omega(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d x\right]\widetilde{\mathbf{z}},$$

where π *is a constant.*

9. *The matching probabilities for sellers and buyers of d-type patents are constant and implicitly defined by*

$$m_a(\frac{n_a}{n_b}) = \eta\left\{\frac{\{1-\sigma[1-m_a(\frac{n_a}{n_b})(1-x_a)]\}(1-\mathbf{i})}{\sigma\mathbf{i}X(x_k)}\right\}^{1-\mu}, \tag{27}$$

and

$$m_b(\frac{n_a}{n_b}) = \eta \left\{ \frac{\sigma \mathbf{i} X(x_k)}{\{1-\sigma[1-m_a(\frac{n_a}{n_b})(1-x_a)]\}(1-\mathbf{i})} \right\}^{\mu}. \tag{28}$$

10. *The constants* $\mathfrak{a}$, $\mathfrak{b}_1$, $\mathfrak{b}_2$, $\mathfrak{k}_1$, $\mathfrak{k}_2$, π, $\mathfrak{s}_1$, $\mathfrak{s}_2$, $\mathfrak{v}_1$, $\mathfrak{v}_2$, x_a *and* x_k *are determined by a nonlinear equation system, in conjunction with the 5 equations (22), (25), (26), (27) and (28) that determine the 5 variables* $\mathbf{g}, \mathbf{i}, r, m_a(n_a/n_b)$, *and* $m_b(n_a/n_b)$. *This system of nonlinear equations does not involve either* $\widetilde{z}$ *or* $\widetilde{\mathbf{z}}$.

Along a balanced growth path, wages grow at the constant gross rate $\mathbf{g}^{\zeta/(\zeta+\lambda)}$, a fact evident from equation (20). So will aggregate output and profits, as can be seen from (7). The gross interest rate, $1/r$, will remain constant along balanced growth. Point 2 implies that on average the values of the firm at the buying, selling, and keeping stages also grow at the rate of growth of output. So, the relative values of a firm at these stages remain constant along a balanced growth path. Thus, it is not surprising then that the decisions to buy, sell or keep d-type patents in terms of propinquity, x, do not change over time. Hence, the function $I_k(z, x; \mathbf{z})$ does not depend on $\mathbf{z}$. It may seem surprising that the decision doesn't depend on z, either. This transpires because a firm's profits are linear in z, as (7) shows. It turns out that $\mathfrak{k}_1 = \mathfrak{s}_1$, which implies that only x is relevant [when comparing $\mathfrak{k}_1\widetilde{z}+\mathfrak{k}_2(x)\widetilde{\mathbf{z}}$ with $\mathfrak{s}_1\widetilde{z}+\mathfrak{s}_2\widetilde{\mathbf{z}}$]. Likewise, the value of a patent agent also increases at rate $\mathbf{g}^{\zeta/(\zeta+\lambda)}$–point 3. Hence, equation (21) dictates that the price, q, at which a firm can sell a d-type patent must also grow at this rate. Additionally, it is easy to see from (16) that the price at which the agent sells a d-type patent to firms, p, will appreciate at this rate too. Note that this price does not depend on z, because given the linear form of the value function, V, only x will be relevant (when comparing $\mathfrak{v}_1 z'$ with $\mathfrak{v}_1 z$). Additionally, using (17) it should now be not too difficult to see that the function $I_a(z, x; \mathbf{z})$ will only depend on x. It's easy to deduce from equation (14) that the rate of innovation, $\mathbf{i}$, will be constant over time if B, K, and S grow at the same rate as aggregate productivity. Since the decisions to buy and sell patents only depend on x, it is straightforward that the number of buyers and sellers on the patent market are fixed along a balanced growth path. To see that the form for the distribution function over buyers, $G(z)$, does not matter note that this function only enters value function for the patent agent (15). But, by points (4) and (8) the functions $I_a(z, x; \mathbf{z})$ and $P(z, x; \mathbf{z})$ do not depend on z. Thus, $G(z)$ is irrelevant in (15). Last, the evolution of shape of the distribution function Z over time does not matter for the analysis. Its mean grows at the gross rate $\mathbf{g}$, independently of any transformation in shape.

Proof of the Existence of a Balanced Growth Path. The proof proceeds using a guess and verify procedure (or the method of undetermined coefficients).

Point (1). To derive the interest and rental rates, imagine the problem of a consumer/worker who can invest in one period bonds that pay a gross interest rate of $1/r$. The Euler equation for asset accumulation will read

$$c^{-\varepsilon} = (\beta/r)(c')^{-\varepsilon}.$$

Along a balanced growth path, if the mean level of productivity grows at rate $\mathbf{g}$ then consumption, the capital stock and output must grow at rate $g^{\zeta/(\zeta+\lambda)}$. This fact can be gleaned from the production function (1), by assuming z grows at rate g, that capital and output grow at another common rate, and that labor remains constant. Therefore, $r = \beta/\mathbf{g}^{\varepsilon\zeta/(\zeta+\lambda)}$. In standard fashion, the rental rate on capital is given by $\widetilde{r} = 1/r - 1 + \delta = \mathbf{g}^{\varepsilon\zeta/(\zeta+\lambda)}/\beta - 1 + \delta$.

Point (4). The form of the threshold rule for buying a d-type patent follows from the fact the sum of the surplus (sans price) accruing to a firm that buys a patent and the surplus (sans price) to the patent agent must be greater than zero; otherwise, a non-negative sale price, p, for the d-type patent would not exist. First, plug the solutions for w and $\widetilde{r}$, or (20) and (23), into the profit function (7) to obtain

$$e\Pi(z, \mathbf{z}) = \pi \frac{ez}{\mathbf{z}^{\lambda/(\zeta+\lambda)}} = \pi e\tilde{z}, \tag{29}$$

and

$$E[e\Pi(z, \mathbf{z})] = \pi\tilde{z}, \text{since } E[e] = 1,$$

with

$$\pi \equiv \frac{\zeta}{\mathbf{g}^{\lambda/(\zeta+\lambda)}} \left(\frac{\kappa}{\mathbf{g}^{\varepsilon\zeta/(\zeta+\lambda)}/\beta + \delta - 1} \right)^{\kappa/(\zeta+\lambda)}. \tag{30}$$

Second, conjecture that the value functions $V(z; s)$ and $A(s)$ have the forms $V(z; s) = \mathfrak{v}_1\tilde{z} + \mathfrak{v}_2\widetilde{\mathbf{z}}$ and $A(s) = \mathfrak{a}\widetilde{\mathbf{z}}$. Third, given the above, note that the (sans price) surpluses for a buying firm and the patent agent are given by

$$\pi(\tilde{z} + \gamma_d x\widetilde{\mathbf{z}}) - \pi\tilde{z} + rE[V\left(z + \gamma_d x\mathbf{z}, \mathbf{z}'\right)] - rE[V\left(z, \mathbf{z}'\right)] = (\pi + \frac{r\mathfrak{v}_1}{\mathbf{g}^{\lambda/(\zeta+\lambda)}})\gamma_d x\widetilde{\mathbf{z}},$$

and

$$-\sigma rA(\mathbf{z}') = -\sigma r\mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a}\widetilde{\mathbf{z}} \text{ [cf. (17)]}.$$

It is easy to deduce from (16) and (17) that sum of these two quantities must be positive for a trade to take place. Note that whether or not the sum of the above two equations is nonnegative does not depend on $\widetilde{\mathbf{z}}$. This sum is also increasing in x. Solving for the value of x that sets the sum to zero yields

$$x_a = \frac{\sigma r\mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a}}{(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d}. \tag{31}$$

Thus, x_a is a constant.

Point (8). The solutions for d-type patent prices are easy to obtain. Insert the above formulae for the (sans price) surplus for a buying firm and the (sans price) surplus for a patent agent into expression (16) to get

$$P(\mathbf{z}, x; \mathbf{z}) = \Big[(1-\omega)\sigma r \mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a} + \omega(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d x\Big]\widetilde{\mathbf{z}}.$$

It is immediate from (21) that $q = \mathfrak{a}\widetilde{\mathbf{z}}$, predicated upon the guess $A(s) = \mathfrak{a}\widetilde{\mathbf{z}}$.

Point (5). It will now be shown that the value function for the patent agent has the conjectured linear form. Focus on equation (15), which specifies the solution for A. The price for a d-type patent does not depend on z, given Point (8). Additionally, $D(x) = U[0,1]$. Furthermore, $I_a = 1$ for $x > x_a$ and is zero otherwise. Thus,

$$A(\mathbf{z}) = \mathfrak{a}\widetilde{\mathbf{z}} = m_a(n_a/n_b)\int_{x_a}^{1} P(\mathbf{z}, x; \mathbf{z})dx + [1 - m_a(n_a/n_b)Pr(x \geq x_a)]\sigma r A(\mathbf{z}'),$$

from which it follows that

$$\begin{aligned}\mathfrak{a} \;=\;& \sigma r \mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a} - m_a(n_a/n_b)(1-x_a)\omega\sigma r\mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a} \\ &+ m_a(n_a/n_b)\omega(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d(1-x_a)(1+x_a)/2.\end{aligned} \tag{32}$$

Point (2). The value function for a buying firm can be determined in a manner similar to that for A in Point (5). Here

$$B(z, \mathbf{z}) = \mathfrak{b}_1\tilde{z} + \mathfrak{b}_2\widetilde{\mathbf{z}},$$

with

$$\mathfrak{b}_1 = \pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)}, \tag{33}$$

and

$$\begin{aligned}\mathfrak{b}_2 \;=\;& -m_b(n_a/n_b)(1-x_a)(1-\omega)\sigma r\mathbf{g}^{\zeta/(\zeta+\lambda)}\mathfrak{a} + r\mathfrak{v}_2\mathbf{g}^{\zeta/(\zeta+\lambda)} \\ &+ m_b(n_a/n_b)(1-\omega)(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d(1-x_a)(1+x_a)/2 \\ &+ (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\mathfrak{p}.\end{aligned} \tag{34}$$

To derive this solution, the results in Points (4) and (8), along with the conjectured solution for V, are used in equation (8). Similarly, using equation (11) it can be shown that the value function for a seller is given by

$$S(z, \mathbf{z}) = \mathfrak{s}_1\tilde{z} + \mathfrak{s}_2\widetilde{\mathbf{z}},$$

with

$$\mathfrak{s}_1 = \pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)}, \tag{35}$$

and

$$\mathfrak{s}_2 = \sigma\mathfrak{a} + r\mathfrak{v}_2\mathbf{g}^{\zeta/(\zeta+\lambda)} + (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\mathfrak{p}. \tag{36}$$

Last, following from (10),

$$K(z + \gamma_d x\mathbf{z}; \mathbf{z}) = \mathfrak{k}_1\tilde{z} + \mathfrak{k}_2(x)\widetilde{\mathbf{z}},$$

with

$$\mathfrak{k}_1 = \pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)}, \tag{37}$$

and

$$\mathfrak{k}_2(x) = (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d x + r\mathfrak{v}_2\mathbf{g}^{\zeta/(\zeta+\lambda)} + (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\mathfrak{p}. \tag{38}$$

Point (3). The threshold rule for keeping or selling a d-type patent is determined by the condition

$$\mathfrak{k}_1\tilde{z} + \mathfrak{k}_2(x_k)\widetilde{\mathbf{z}} = \mathfrak{s}_1\tilde{z} + \mathfrak{s}_2\widetilde{\mathbf{z}};$$

that is, at the threshold a firm is indifferent between keeping or selling the patent. Now, $\mathfrak{s}_1 = \mathfrak{k}_1$ so that

$$(\pi+r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_d x_k+r\mathfrak{v}_2\mathbf{g}^{\zeta/(\zeta+\lambda)}+(\pi+r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\mathfrak{p} = \sigma\mathfrak{a}+r\mathfrak{v}_2\mathbf{g}^{\zeta/(\zeta+\lambda)}+(\pi+r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\mathfrak{p}.$$

Hence,

$$x_k = \frac{\sigma\mathfrak{a}}{\left[\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)}\right]\gamma_d}, \tag{39}$$

a constant.

Point (6). Turn now to the beginning-of-period value function for the firm and the rate of innovation that it will choose. By using the linear forms for the value functions $B(z, z)$, $K(z+\gamma_d xz; s)$, and $S(z, z)$, the fact that $\mathfrak{b}_1 = \mathfrak{k}_1 = \mathfrak{s}_1$, and the property that the threshold rule takes the form $I_k = 1$ for $x > x_k$ and $I_k = 0$ otherwise, the firm's dynamic programming problem (13) can be rewritten as

$$V(z, \mathbf{z}) = \widetilde{\mathbf{z}} \max_{i\in[0,1]}\{[X(x_k)\mathfrak{s}_2 + \int_{x_k}^{1} \mathfrak{k}_2(x)dX(x) - \mathfrak{b}_2]i - \frac{\chi}{1+\rho}i^{1+\rho}\} + (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\tilde{z} + \mathfrak{b}_2\widetilde{\mathbf{z}}.$$

Differentiating with respect to i then gives

$$X(x_k)\mathfrak{s}_2 + \int_{x_k}^{1} \mathfrak{k}_2(x)dX(x) - \mathfrak{b}_2 = \chi i^{\rho},$$

from which (25) follows. Using the solution for i, as given by (25), in the above programming problem yields

$$V(z, \mathbf{z}) = \frac{\rho}{(1+\rho)\chi^{1/\rho}}[X(x_k)\mathfrak{s}_2 + \int_{x_k}^{1} \mathfrak{k}_2(x)dX(x) - \mathfrak{b}_2]^{1+\frac{1}{\rho}}\widetilde{\mathbf{z}} + (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\tilde{z} + \mathfrak{b}_2\widetilde{\mathbf{z}}.$$

It then follows that

$$\mathfrak{v}_1 = \frac{\mathbf{g}^{\lambda/(\zeta+\lambda)}}{\mathbf{g}^{\lambda/(\zeta+\lambda)} - r}\pi, \tag{40}$$

and

$$\mathfrak{v}_2 = \mathfrak{b}_2 + \frac{\rho}{(1+\rho)\chi^{1/\rho}}[X(x_k)\mathfrak{s}_2 + \int_{x_k}^{1} \mathfrak{k}_2(x)dX(x) - \mathfrak{b}_2]^{1+\frac{1}{\rho}}. \tag{41}$$

Point (7). The gross rate of growth for aggregate productivity, $\mathbf{g}$, will now be calculated. Suppose that aggregate productivity is currently z. A fraction $\mathbf{i}$ of firms will innovate today. Those firms that draw $x > x_k$ will keep their good patent. The productivity for these firms will increase. The fraction $1 - \mathbf{i}$ of firms will fail to innovate. Out of these firms the proportion $m_b(n_a/n_b)$ will find a seller on the market for d-type patents. They will buy a d-type patent when $x > x_a$. Thus, z will evolve according to

$$\mathbf{z}' = \mathbf{z} + \mathbf{i}\int_{x_k}^{1} \gamma_d x\mathbf{z}dX(x) + m_b(n_a/n_b)(1-\mathbf{i})\int_{x_a}^{1} \gamma_d x\mathbf{z}dx + \gamma_n \mathfrak{p}\mathbf{z}.$$

This implies (26).

Point (9). The number of buyers on the market for d-type patents is given by $n_b = 1-\mathbf{i}$; all firms that fail to innovate will try to buy a d-type patent. Along a balanced growth path, the number of patent agents, n_a, must satisfy the equation

$$n_a = \sigma n_a[1 - m_a(n_a/n_b)(1-x_a)] + \sigma\mathbf{i}X(x_k).$$

Focus on the right-hand side. Take the first term. Suppose that there are n_a patent agents at the beginning of the period. A fraction σ of these agents will survive. Out of these, $m_b(n_a/n_b)(1 - x_a)$ will find a buyer. Thus, they will not be around these next period. Move to the second term. A mass of $\mathbf{i}X(x_k)$ new firms will decide to sell their patents Out of this σ will survive. The sum of these two terms equals the new stock of patent for sale, n_a. Solving yields

$$n_a = \frac{\sigma\mathbf{i}X(x_k)}{1-\sigma[1 - m_a(n_a/n_b)(1-x_a)]} \text{ and } \frac{n_a}{n_b} = \frac{\sigma\mathbf{i}X(x_k)}{(1-\mathbf{i})\{1-\sigma[1 - m_a(n_a/n_b)(1-x_a)]\}}.$$

Equations (27) and (28) follow immediately.

Point (10). The 12 constants, viz a, $\mathfrak{b}_1$, $\mathfrak{b}_2$, $\mathfrak{k}_1$, $\mathfrak{k}_2$, π, $\mathfrak{s}_1$, $\mathfrak{s}_2$, $\mathfrak{v}_1$, $\mathfrak{v}_2$, x_a and x_k, are specified by the 12 non-linear equation (30) to (41). The equations include the variables $\mathbf{g}, \mathbf{i}, r, m_a(n_a/n_b)$, and $m_b(n_a/n_b)$. So, equations (22), (25), (26), (27) and (28) must be appended to the system to obtain a system of 17 equations in 17 unknowns. This system does not depend on either $\widetilde{z}$ or $\widetilde{\mathbf{z}}$. ■

8.2 More on Tacking on a Market for n-type Patents

The discussion in Section 3.4 is completed here. An n-type idea is worth $(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\widetilde{\mathbf{z}}$ in production value to a firm.[22] Specifically, it will increase z' by $\gamma_n\mathbf{z}$. This will lead to increase in current profits in the amount $\pi\gamma_n\widetilde{\mathbf{z}}$ and discounted expected future profits by $r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)}\gamma_n\widetilde{\mathbf{z}}$. Any price, q_b, in the interval $[0, (\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\widetilde{\mathbf{z}}]$ can be an equilibrium market price on the market for n-type patents. The exact value for q_b doesn't matter though. At the time of all decision making, the expected discounted present value of profits arising from an n-type patent is $\mathfrak{p}[(1-\mathfrak{p}_s)(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\widetilde{\mathbf{z}}+\mathfrak{p}_s q_b]+(1-\mathfrak{p})\mathfrak{p}_b[(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\widetilde{\mathbf{z}} - q_b]$, which takes into account the keeping, selling and buying events, respectively. This expression reduces to $\mathfrak{p}(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\widetilde{\mathbf{z}}$, using the fact that $\mathfrak{p}\mathfrak{p}_s = (1-\mathfrak{p})\mathfrak{p}_b$. Thus, the expected discounted present value of profits associated with an n-type patent does not involve the equilibrium market price, q_b, or the buying and selling probabilities, $\mathfrak{p}_b$ and $\mathfrak{p}_s$. Therefore, adding a market for n-type patents does not alter the solution for the balanced growth path presented in Proposition 1.

9 Empirical Appendix

The brunt of the analysis relies on data from three sources: USPTO, NBER Patent Database Project (PDP), and Compustat. The first sources contains data on the patents that are reassigned across firms. The second is used to retrieve information on the technological classes for patents and the stocks of patents for firms. Facts about employments, stock market values for firms are obtained from the third source.

9.1 Patent Reassignment Data (PRD)

The patent assignment data is obtained from the publicly available United States Patent and Trademark Office (USPTO) patent assignment files hosted by Google Patents Beta. These files contain all records of changes made to U.S. patents for the years 1980-2011. The files are parsed and combined to create the dataset. The following variables are kept:

- Patent number: The unique patent number assigned to each patent granted by the USPTO.

[22] In Section 8.1 it is shown that the value functions for buying, keeping, selling and innovating firms are linear in the expected value of a new n-type idea, as can be seen by examining the coefficients, $\mathfrak{b}_2$,$\mathfrak{k}_2$, $\mathfrak{s}_2$, and $\mathfrak{v}_2$. The terms in question have the all form $(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n\mathfrak{p}$, implying that the production value of an n-type idea is $(\pi + r\mathfrak{v}_1/\mathbf{g}^{\lambda/(\zeta+\lambda)})\gamma_n$ – see (34), (36), (38) and (41).

- Record date: Date of creation for the record.
- Execution date: Date for the legal execution of the record.
- Conveyance text: A text variable describing the reason for the creation of the record. Examples are: "Assignment of assignor's interest", "Security Agreement", "Merger", etc.
- Assignee: The name of the entity assigning the patent (i.e., the seller if the patent is being sold).
- Assignor: The name of the entity to which the patent is being assigned (i.e., the buyer if the patent is being sold).
- Patent application date: Date of application for the patent.
- Patent grant date: Date of grant for the patent.
- Patent technology class: The technology class assigned to the patent by the USPTO according to its internal classification system.[23]

The entries for which this information is inaccessible are dropped from the sample.

During the parsing process, the following are done:

- Only transfer agreements between companies are kept.
- Only utility patents are kept; entries regarding design patents are dropped.

This cleaning process leaves 966,427 observations. Using the string variable that states the reason for the record, all reassignments that are not directly related to sales are dropped (for instance, mergers, license grants, splits, mortgages, court orders, etc.)

In order to create an even more conservative indicator of patent reassignments, a company name-matching algorithm is employed, so that marking internal transfers as reassignments can be avoided, where patents are moved within the same firm, or between the subsidiaries of the firm. The idea behind the company name-matching algorithm is to clean the string variables for the assignor and the assignee of all unnecessary indicators and company type abbreviations. If the cleaned assignor and assignee strings are equal, the type of the record is changed to internal transfer, provided that it was identified as a reassignment before.

The pseudo-code for the algorithm, an enhanced version of Kerr and Fu (2008), is as follows:

[23] This variable is not used, however, to represent the technology class for a patent, as is discussed below.

1. All letters are made upper case.

2. The portion of the string after the first comma is deleted. (e.g., AMF INCORPORATED, A CORP OF N.J. becomes AMF INCORPORATED)

3. If the string starts with "THE ", the first 4 characters are deleted.

4. All non-alphanumeric characters are removed.

5. Trailing company identifiers are deleted if found. The string goes through this process 5 times. The company identifiers are the following: B, AG, BV, CENTER, CO, COMPANY, COMPANIES, CORP, CORPORATION, DIV, GMBH, GROUP, INC, INCORPORATED, KG, LC, LIMITED, LIMITED PARTNERSHIP, LLC, LP, LTD NV, PLC, SA, SARL, SNC, SPA, SRL, TRUST, USA, KABUSHIKI, KAISHA, AKTIENGESELLSCHAFT, AKTIEBOLAG, SE, CORPORATIN, CORPORATON, TRUST, GROUP, GRP, HLDGS, HOLDINGS, COMM, INDS, HLDG, TECH, and GAISHA.

6. If the resulting string has length zero, that string is declared as needing protection. Some examples that are protected by this procedure: "CORPORATION, ORACLE", "KAISHA, ASAHI KAISEI KABUSHIKI", "LIMITED, ZELLWEGER ANALYTICS".

7. The algorithm is re-run from the beginning on the original strings with one difference: The strings that are declared as needing protection skip the second step.

9.2 USPTO Utility Patents Grant Data (PDP)

The patent grant data comes from the NBER Patent Database Project (PDP), and contains data for all the utility patents granted between the years 1976-2006. How the PDP and PRD are linked to each other is discussed later on.

9.3 Compustat North American Fundamentals (Annual)

The Compustat data for publicly traded firms in North America between the years 1974-2006 is retrieved from Wharton Research Data Services. The Compustat database and the NBER PDP database are connected using the matching procedure provided alongside the PDP data. Extensive information on how the matching is done can be found on the project website.

9.4 Connecting PRD and PDP Data

There are two different questions of interest, which require combining the Patent Database Project data with the Patent Reassignment Data. The first analysis is on whether a patent is ever reassigned (i.e. sold) over its entire lifetime, and what determines the probability of this event. For this purpose, it is only necessary to connect the information from PRD to the firm which applied for the patent. This is easily done by using the unique patent number each patent is given by USPTO.

The second question involves the change in match quality of the patent when it is traded between two firms. In this case, one needs to know the characteristics of both the assignor and the assignee firm for each reassignment record in the PRD dataset. However, there is no existing connection established between the PRD and PDP datasets. To connect these datasets, the company name-matching algorithm described earlier is employed.

9.5 Variable Construction

9.5.1 Patent-to-Patent Distance Metric

In order to construct a topology on the technology space empirically, it is necessary to create a distance metric between different technology classes. Such a metric enables one to speak about the distance between two patents in the technology space, and leads to the construction of more advanced metrics.

The first 2 digits of the IPC (International Patent Classification) codes of a patent are chosen to indicate its technology class. The IPC code used for a patent is taken from the PDP data and differs from the classification scheme employed in the PRD data. It should be noted that the PDP data actually contain more than a single IPC class for a single patent in some cases, since the IPC codes were assigned using a concordance between the IPC and the internal classification system of USPTO. The IPC code provided in the PDP file with assignees is used in such cases, which is unique for each patent.

As discussed in the main text, a plausible distance metric between patent classes can be generated by looking at how often two different technology classes are cited together. Formally:

$$d(X,Y) \equiv 1 - \frac{\#(X \cap Y)}{\#(X \cup Y)}, \text{ with } 0 \leq d(X,Y) \leq 1.$$

where $\#(X \cap Y)$ denotes the number of patents that cite technology classes X and Y together, and $\#(X \cup Y)$ denotes the number of patents which either cite X or Y or both.

9.5.2 Definition of a Firm in the Data

There are four different entity identifiers in the PDP dataset. The USPTO assignee number is the identifier provided by USPTO itself, but the creators of the PDP have found that it is not very accurate. A single assignee might have many different USPTO assignee numbers. The PDP uses some matching algorithms on the names of the assignees to create a more accurate assignee identifier, called PDPASS. They also link the patent data to Compustat data. Compustat has an identifier called GVKEY. However, these refer to securities rather than firms. So a single firm might be represented by many GVKEY's. For this reason, they use a dynamic matching algorithm again to link all GVKEY's to certain PDPCO's, where the latter is a unique firm identifier that is created by the authors of the project. They create this identifier in order to be able to account for name changes, mergers & acquisitions, etc. This paper follows the same procedure.

9.5.3 Patent-to-Firm Distance Metric

In order to measure how close a patent is to a firm in the technology spectrum, a metric is necessary. However, throughout their lifetimes firms register patents in multiple technology classes. Hence the patent-to-patent distance metric is insufficient for this purpose. One possible way to construct a patent-to-firm distance metric is to compare a patent to the past patent portfolio of the firm. The distance measure between each patent a firm registered in the past, and the new patent in question can be calculated using the patent-to-patent distance metric offered earlier. The distance between the firm and the patent should be a function of these separate distances. Equation (24) defines a parametric family of distance measures indexed by ι. The analysis is conducted for several values of ι.

9.5.4 Creating the Patent Stock Variable for Compustat Firms

As argued in Hall et al (2005), the citation-weighted patent portfolio of a firm is a plausible indicator of the intangible knowledge stock of a firm. The authors demonstrate that this measure has additional explanatory power for the market value of a firm above and beyond the conventional discounted sum of R&D spending of a firm, since R&D is a stochastic process which can succeed or fail; whereas patents are quantifiable products of this process when it is successful. Furthermore, it is revealed that the number of citations a patent receives is a fine indicator of the patent's worth, increasing the market value of a firm at an increasing rate as the number of citations go higher.

Since all the future citations to a patent cannot be observed at any given date, the citations

Patent Sale Decision
(Compustat Sample with Licensing Intensity)

	Dependent Variable (=1, if sold, 0 otherwise)					
	(1)	(2)	(3)	(4)	(5)	(6)
distance	3.981***	3.981***	3.994***	4.351***	4.350***	4.677***
	(0.138)	(0.138)	(0.138)	(0.142)	(0.142)	(0.157)
tech-class litigation intensity	no	yes	no	no	yes	no
patent litigation dummy	no	no	yes	no	yes	no
sector licensing intensity	no	no	no	yes	yes	no
only renewed patents	no	no	no	no	no	yes
Obs	1,162,182	1,162,182	1,162,182	1,088,874	1,088,874	928,169
R^2	0.32	0.32	0.32	0.32	0.32	0.34

Table 8: See the notes for Table 3.

variable suffers from a truncation problem. There are also technology class and year fixed effects to consider. All of these issues are thoroughly investigated by Hall et al (2005); they provide a variable called hjtwt in order to correct the citation number of each patent in the PDP dataset. This study uses their correction method. In the end, a corrected citations number for each patent is obtained. In order to create the patent stock variable of a firm (PDPCO), the corrected citations number across all the patents of a firm are added together at each year. This variable is called patent stock.

In addition to the patent stock, the corrected citations number across all the patents of a firm, multiplied by the patent-to-firm distance generated at the date of the patent's inclusion into the portfolio are also added together to create a new variable. This variable quantifies the overall waste in the patent stock caused by the mismatch between the technology class of the patents and the firm. This variable is expected to have a negative effect on firm moments such as real sales and market value of equity. The variable is called the distance-adjusted patent stock.

9.6 Patent Sale Decision with Licensing Intensity

Table 8 introduces the licensing intensity of the sector. This variable is available only for Compustat firms. Therefore the sample is reduced by half. Because of this sizable change, columns 1-3 repeat the same exercises as their counterparts in Table 3. One major difference to note is that the association between the distance and sale indicators becomes more pronounced, almost doubled. Column 4 introduces licensing intensity and column 5 introduces all litigation and licensing controls simultaneously. The last column redoes the regression in column 1 while purging the patents that were not renewed once.

Calibration/Estimation Jacobian (Elasticities, %)

Param	Growth	R&D/GDP	Frac. sold	Avg. Dur.	Dur. CV	daps/ps	ps/emp	dist red, all
γ_d	74.39	47.39	-8.64	0.36	-0.18	15.71	-4.64	-4.12
χ	-17.18	-10.55	23.06	-4.81	2.42	-7.68	-23.37	4.67
μ	0.86	-1.41	9.63	-4.22	2.12	-4.27	-2.03	18.87
η	3.32	-5.45	37.25	-16.34	8.21	13.00	-8.60	73.06
γ_n	22.46	-24.81	5.66	-1.75	0.88	-17.79	-98.84	0.21
$\mathfrak{p}$	22.46	-24.81	47.46	-1.75	0.88	-3.74	-51.07	-71.53
$\mathfrak{p}_s$	0	0	64.57	0	0	18.55	-36.70	-71.74
STD($\ln e$)	0	0	0	0	0	-3.84	225.95	0

Table 9: The data targets in the Jacobian follow the order that they are presented in Table 5.

9.7 The Impact of Parameter Values on the Data Targets

Table 9 presents the Jacobian associated with the calibration/estimation. This Jacobian provides useful information about how the parameters influence the model's ability to hit the data targets. By moving along a row, one can see how a parameter in question influences the various data targets. Alternatively, by going down a column one can gauge what parameters are important for hitting the data target of concern.

SUPPLEMENTAL APPENDIX
(For Working Paper Only)

10 Background on the Market for Patents–Supplemental Material

In this section additional background information on the market for patents is provided. The presentation starts with some historical evidence in Section 10.1 and then turns to more recent evidence in Section 10.2.

10.1 Historical Evidence

Patents constitute a property right over a technology the ownership of which can be transferred to a secondary party. This right was enacted by the Patent Act of 1836, which states:

> "And be it further enacted, That every patent shall be assignable in law, either as to the whole interest, or any undivided part thereof, by any instrument in writing; which assignment, and also every grant and conveyance of the exclusive right under any patent, to make and use, and to grant to others to make and use, the thing patented within and throughout any specified part or portion of the United States, shall be recorded in the Patent Office within three months from the execution thereof, for which the assignee or grantee shall pay to the Commissioner the sum of three dollars."

Since the 1836 Act, the U.S. Patent Office has recorded information on the ownership of all U.S. patents. Many patents have been sold in the market. Unfortunately, despite its importance and wide use, the empirical studies on the market have been limited due to the lack of systematic data [Lamoreaux, Sokoloff, and Sutthiphisal (2013)]. Recently, however, Khan and Sokoloff (2004) have gone through the *Dictionary of American Biography* and collected systematic data on "great inventors" of the 19th century. Their findings are striking: For inventors who started their own firms, it is seen that only one third of the patents in their patent portfolio were actually granted to them, which implies the remaining two thirds were acquired from others. Their study documents the size of the market. The fraction of patents that had a reassignment was between 16-44% during the second half of the 19th century. This is very similar to the number found here that lies between 14-22% for the period 1980-2000.

Khan (2013) argues that the market for patents in the U.S. developed very rapidly thanks to the effective patent and legal system in the country. She (p. 8) states "As a result, American inventors were able to benefit from patent markets to a far greater extent than in other countries. Intermediation enhanced their ability to divide and subdivide the rights to their idea, sometimes with great complexity, across firms, industries and regions. Successful inventors were able to leverage their reputations and underwrite the research and development costs of their inventions by offering shares in future patents. This process also facilitated trade in patent rights and technological innovations across countries, and numerous American patentees succeeded in establishing multinational enterprises and dominating the global industry."

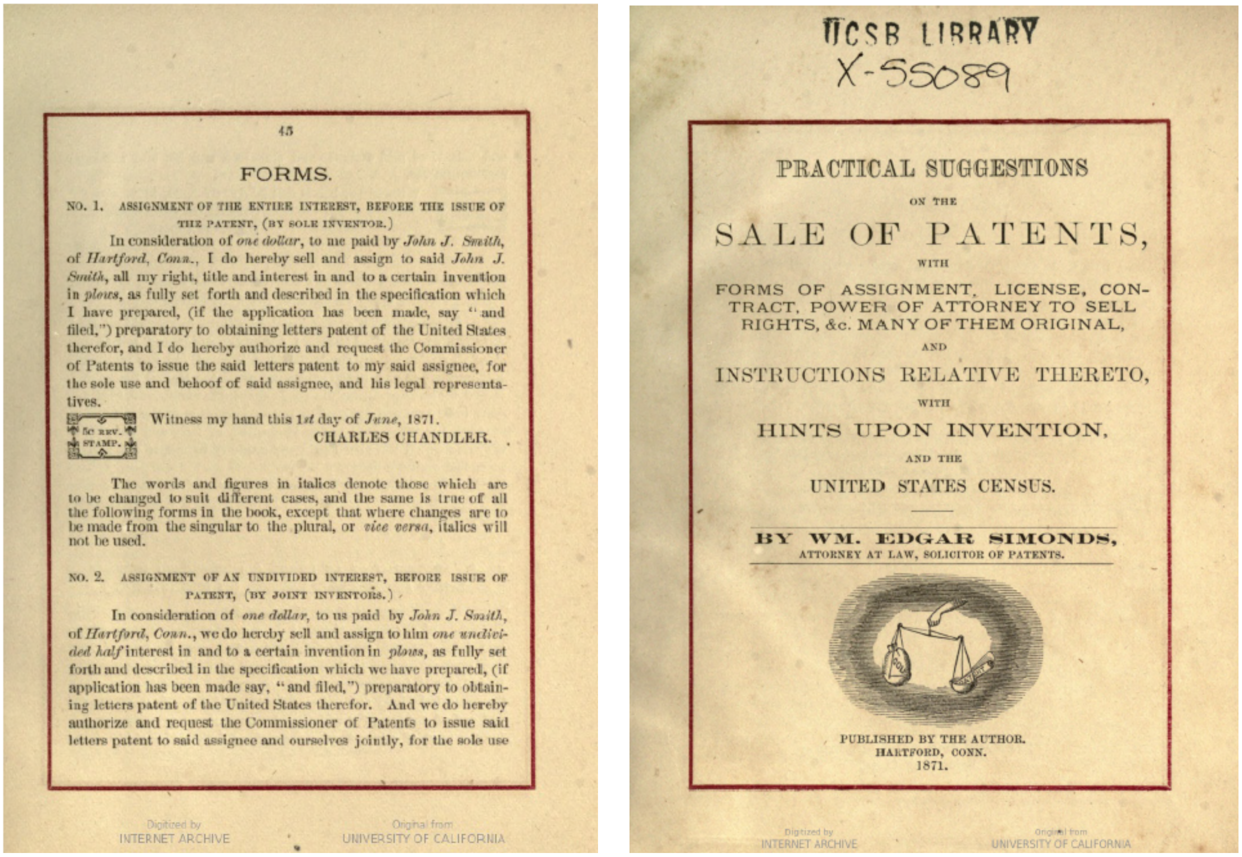
This flexible environment and the possibility of selling their inventions provided many inventors, with great potential, the chance to flourish. For instance, Thomas Edison transferred the partial rights for 20 of the first 25 patents in his career [Lamoreaux, Sokoloff, and Sutthiphisal (2013)]. Overall, the existence of the market allowed the *democratization* of innovation. It provided small-scale garage inventors with access to the market for technology. The same is true for female inventors and Khan (2013) provides various examples on how they benefited, in particular, from the market. For instance, Maria Beasley reached an agreement in 1881 to transfer half of the rights for an uncompleted invention to James Henry of Philadelphia, in return for an advance of funds to complete the machine.

The market has been key for allocating important innovations to the right hands. For instance, Nicholas (2009) uses geo-coded data on the location of inventors and research labs to show that a significant fraction of the most valuable patents acquired by firms during the 1920s were most likely not generated in the firms' research laboratories.

10.1.1 Intermediaries

Many inventors who tried to sell their inventions in the market, such as Rufus M. Porter who invented the alarm clock, washing machine, clothes dryer, and rotary plow, failed because they could not find interested assignees. This market has been associated with severe matching frictions. Kahn (2013) argues that intermediaries have the ability to reduce the costs of search and exchange, to enhance liquidity, to improve market depth and breadth, and to increase overall efficiency. Lamoreaux and Sokoloff (1999) report that there were 550 such registered patent agents by 1880. Figure 8 depicts an example of a contract that was prepared around 1870 to transfer to the ownership rights of a patent.

Even during the 19th century, there were manuals for inventors which taught how to sell their patents more easily. Figure 8 shows the cover of one such manual prepared by William E. Simonds in 1871. This manual was advising inventors to advertise their inventions as much as possible. Advertising patents was key for finding a buyer. For instance, Elias E. Reis read an advertisement



45

FORMS.

NO. 1. ASSIGNMENT OF THE ENTIRE INTEREST, BEFORE THE ISSUE OF THE PATENT, (BY SOLE INVENTOR.)

In consideration of *one dollar*, to me paid by *John J. Smith*, of *Hartford, Conn.*, I do hereby sell and assign to said *John J. Smith*, all my right, title and interest in and to a certain invention in *plows*, as fully set forth and described in the specification which I have prepared, (if the application has been made, say "and filed,") preparatory to obtaining letters patent of the United States therefor, and I do hereby authorize and request the Commissioner of Patents to issue the said letters patent to my said assignee, for the sole use and behoof of said assignee, and his legal representatives.

5c REV. STAMP.

Witness my hand this *1st* day of *June*, 1871.

CHARLES CHANDLER.

The words and figures in italics denote those which are to be changed to suit different cases, and the same is true of all the following forms in the book, except that where changes are to be made from the singular to the plural, or *vice versa*, italics will not be used.

NO. 2. ASSIGNMENT OF AN UNDIVIDED INTEREST, BEFORE ISSUE OF PATENT, (BY JOINT INVENTORS.)

In consideration of *one dollar*, to us paid by *John J. Smith*, of *Hartford, Conn.*, we do hereby sell and assign to him *one undivided half* interest in and to a certain invention in *plows*, as fully set forth and described in the specification which we have prepared, (if application has been made say, "and filed,") preparatory to obtaining letters patent of the United States therefor. And we do hereby authorize and request the Commissioner of Patents to issue said letters patent to said assignee and ourselves jointly, for the sole use

PRACTICAL SUGGESTIONS

ON THE

SALE OF PATENTS,

WITH

FORMS OF ASSIGNMENT, LICENSE, CONTRACT, POWER OF ATTORNEY TO SELL RIGHTS, &c. MANY OF THEM ORIGINAL,

AND

INSTRUCTIONS RELATIVE THERETO,

WITH

HINTS UPON INVENTION,

AND THE

UNITED STATES CENSUS.

BY WM. EDGAR SIMONDS,
ATTORNEY AT LAW, SOLICITOR OF PATENTS.

PUBLISHED BY THE AUTHOR.
HARTFORD, CONN.
1871.

Figure 8: Patent Sale Contract around 1870 (left panel) and the Cover of a Patent Sale Manual by Edgar Edgar Simonds, 1871 (right panel).

about a patent on electrical welding, invented by Elihu Thomson in 1886, and said it "immediately opened up to my mind a field of new applications to which I saw I could apply my system of producing heat in large quantities" [Lamoreaux and Sokoloff (1999, p. 23)].

One important observation made by Khan (2013) is as follows. Even though the popular media discusses patent intermediaries [using modern jargon: patent trolls or non-practicing entities (NPE)] as if they have emerged recently, such entities have been the norm in this market throughout the history. NPEs were the norm during the nineteenth century, and technology markets provide ample evidence that patentees who licensed or assigned their rights were typically the most productive and specialized inventors. NPEs did not produce anything with the patents that they purchased. They used their expertise and network to find the right users who could generate the largest expected economic return from these inventions. NPEs profited from intermediation, per se, without participating in either inventive activity or manufacturing.

10.1.2 Size of the Market for Patents

Even in yesteryear the market for patents was sizable [Lamoreaux and Sokoloff (1997, 1999)]. Lamoreaux, Sokoloff, and Sutthiphisal (2013) have shown that the fraction of patents that had a reassignment was somewhere between 16 to 28% in the late 19th century, a number similar to the 14 to 22% found here.

10.2 More Recent Evidence

10.2.1 Size of the Market for Patents

Several papers have documented that the size of the market for technologies in the U.S. and Europe in 1990s were around 0.2% of their GDPs [Arora, Fosfuri, and Gambardella (2001), Gambardella, Giuri, and Mariani (2006)]. In particular, using figures from Internal Revenue Services, Arora et al (2001) estimate the monentary value of the overall volume market for patents in the U.S. to be around $32 billion in 2000. A McKinsey report estimates the same number to be $100 billion in 2000 [Elton et al. (2002)]. In a European Commission (EC) report, Gambardella et al (2006) report that the market for patents in Europe was €9.4 billion between 1994 to 1996 and went up to €15.6 billion in 2000 to 2002, which corresponds to 0.16 and 0.2% of GDP. Similarly, Serrano(2013)'s estimates, based on patent reassignment data, show that the volume of patent sales correspond to 50% of the total valuation of patents.

The same reports highlight the fact that even though many firms wanted to sell their technologies, they could not do so due to market frictions. For instance, Gambardella et al (2006) estimate the potential market sizes for "sleeping patents" in 1994-1996 and 2000-2002 periods as €14.8 billion and €24.4 billion, which are larger than the actual market sizes. Elton et al (2002, p. 2) report that while companies could potentially earn up to 10% of their operating income from utilizing their patents in technology sales, only 0.5% of patents are actually utilized currently. They conclude that this underutilization is mainly driven by managerial failures and informational asymmetries in the market for patents.

The patent market is regarded as being thin due to the specialized nature of the knowledge that is embodied in each patent. Gans and Stern (2010) and Hagiu and Yoffie (2011) discuss the recent failure of online intellectual property platforms to arbitrage the market. According to them, the sensitivity of intellectual property makes potential buyers and sellers reluctant to reveal information online; they prefer face-to-face dealings with the other party. Also, some buyers may perceive a lemons problem: if the patents were truly valuable, then the sellers should be able to profit by developing the idea themselves or by selling it directly to interested parties.

10.2.2 The Complementarity between New Patents and the Existing Stock of Knowledge within a Firm

The significance of the propinquity of a idea with a firm's core line of business is recognized in Gort (1962)–see Jovanovic (1993) for a formalization of some of the ideas in Gort (1962). In well-known work, Teece (1986) stresses the importance of asset complementary as a key ingredient for successful innovation. The closer an innovation is to a firm's core line of business, the more likely it is to have the technical knowledge to implement it and the practical knowledge to market it. Arora and Ceccagnoli (2006) and Figueroa and Serrano (2013) examine the empirical significance of this idea for patenting and licensing activities. Relatedly, several papers have shown that firms' internal R&D activities affect the type of patents that they acquire [Arora and Gambardella (1994) and Cassiman and Veugelers (2006)]. In addition, Salant (1984) provides a theoretical analysis of ideas trading as motivated by preventing the dissipation of competitive rents, a hypothesis examined empirically by Gans and Stern (2000). This can only be successful when the idea is close to the line of business that the firm is trying to protect from competitors.

10.2.3 Serendipity

The European Commission report finds that 1/3 of European patents are not used for any industrial purpose. Why are there so many unused patents and why are they produced in the first place? The report states that half of these unused patents are "sleeping patents" that are typically by-product inventions in non-core technologies for which the inventing firm cannot foresee a potential use. Similarly, Sakkab (2002) provides an interesting case study: Proctor and Gamble (P&G) was commercializing only 10% of its patents and the rest were "sitting on their shelf." Chesbrough (2006) argues that this huge underutilization of ideas in firms is due to the very decentralized process that determines which projects research staff work on and what discoveries are made in the firm. Many firms recruit R&D personnel by promising research freedom and often compete with universities. This process limits the coupling of research ideas with business ideas. Hence, firms can produce ideas that are not close to their core business activities and cannot foresee a potential use of these ideas either. Elton et al (2002, p. 2) say that "Engineers at chemical companies, for example, aren't likely to know that the materials and processes they use to separate atmospheric gases could help semiconductor manufacturers reduce the time and money needed to manufacture the high-value integrated circuits that use ceramic rather than plastic bindings. (Ceramic can withstand more heat than plastic and thus allows for smaller sizes and higher densities.) Yet one midsize chemical company, helped by an external network of technologists, discovered that its process could cut the production costs of these chips by up to 20 percent, or more than $200 million."

10.2.4 Policy Implications

This huge patent market has important policy implications. Some scholars have argued that despite the rapid growth in the potential of the market for ideas, industrial policies are not following this rapid change and hence are creating obstacles limiting better market allocations. Gambardella et al (2006) call policymakers to action in order to increase the rate of utilization of patents. Similarly, Chesbrough (2006) argues that there is no information standard for intellectual property trade. Without these standards, it is very difficult to collect aggregate statistics on this trade and it becomes much harder for firms to know what technologies are available in the market. Indeed, according to a survey by Radauer and Dudenbostel (2013) one of the major obstacles that firms are reporting is the difficulty in identifying suitable partners in the market for ideas. Both economists and policymakers have emphasized the need for revising industrial policies in light of the presence of a market for ideas.[24]

The main policy conclusions of these studies concern reducing informational frictions in the market through better intermediation. For instance, Gambardella et al (2006) advocate policies that would simplify the formation of intermediaries. For interested readers, Tietze (2010) provides further details, as well as a detailed literature review on patent intermediaries.

11 Stationary Firm-Size Distribution–Supplemental Material

The model will display a stationary firm-size distribution along a balanced growth, despite the fact that the distribution for Z is shifting over time and changing shape. To see this, substitute (20) into (19), while making use of the definition in (18), to get

$$l = \frac{e'z'}{\mathbf{z}'} \equiv e'\widehat{z}'. \tag{42}$$

Thus, the amount of labor that a firm hires is proportional to its own productivity, $e'z'$, relative to the mean level of productivity in the economy, $\mathbf{z}'$. The density function for l is the just the

[24] This is in the summary of a European Patent Office-OECD-UK Patent Office Conference, entitled "Patents: Realising and Securing Value" held on November 21, 2006. It can be accessed at: http://www.oecd.org/science/sci-tech/37952293.pdf.

density function for e' times by the density function for z', because e' and z' are independent of each other. The probability distribution function for the idiosyncratic firm-specific production shock, e', is exogenously given. So, characterizing the firm-size distribution amounts to characterizing the distribution for $\widehat{z}' \equiv z'/\mathbf{z}'$.

To delineate the firm-size distribution, focus on a firm's draw for x. This is an independently and identically distributed random variable. To see this, note that in the current setting a firm will discover a d-type idea with probability i. If it innovates then it will draw x from the distribution X. Conditional on innovating, it will sell its d-type patent with probability $X(x_k)$ and will keep it with probability $1 - X(x_k)$. If it fails to innovate then it will go onto the market for d-type patents. Conditional on failing to innovate, it finds a patent agent with probability $m_b(n_a/n_b)$. When it finds a patent agent then it will draw from $\mathcal{U}[0,1]$. A purchase then occurs with probability $1 - x_a$. Hence,

$$x \text{ is } \begin{cases} = 0, & \text{with } \Pr\{(1-i)[(1-m_b) + m_b x_a] + iX(x_k)\}, \\ \sim X[x_k, 1], & \text{with } \Pr[i(1 - X(x_k))], \\ \sim \mathcal{U}[x_a, 1], & \text{with } \Pr[(1-i)m_b(1-x_a)]. \end{cases} \tag{43}$$

Note that $0 < E[x] < 1$.

Turn to the firm's law of motion for z or (2). Divide z through by $\mathbf{z}$ to get

$$\frac{z'}{\mathbf{z}'} = \frac{\mathbf{z}}{\mathbf{z}'}\frac{z}{\mathbf{z}} + \gamma_d \frac{\mathbf{z}}{\mathbf{z}'}x + \gamma_n \frac{\mathbf{z}}{\mathbf{z}'}b, \quad \text{or} \quad \widehat{z}' = \underbrace{\frac{1}{\mathbf{g}}}_{<1}\widehat{z} + \gamma_d \frac{1}{\mathbf{g}}x + \gamma_n \frac{1}{\mathbf{g}}b, \tag{44}$$

where $\widehat{z} \equiv z/\mathbf{z}$. This is a stationary autoregressive process with a non-Gaussian error term. Here, treat $\mathbf{g}$ as a constant, The gross growth rate, $\mathbf{g}$, can be taken as a constant because it can be solved for independently of the form for the stationary distribution. Proposition 1 establishes this. In a similar vein, i, m_b, x_a, and x_k are known constants that are independent of the form of the stationary distribution; again, this is a consequence of Proposition 1. Now, $0 \leq b, x \leq 1$. Thus, the process for $\widehat{z}'$ will be trapped within the compact set $[0, \overline{z}]$, where $\overline{z} \equiv (\gamma_d + \gamma_n)/[(\mathbf{g} - \mathbf{1})]$, provided that it starts off within this interval.

Proposition 2 *(Existence of a Unique Stationary Firm-Size Distribution). The stochastic process (44) converges weakly to a unique invariant distribution.*

Denote the stationary distribution for $\widehat{z}'$ by $\widehat{Z}$. Why does the stationary firm-size distribution (modulo the part associated with the distribution for e) have a finite upper bound, $\overline{z}$? The answer is that it is difficult for a firm's productivity, z, to grow faster than aggregate productivity, $\mathbf{z}$. Growth in

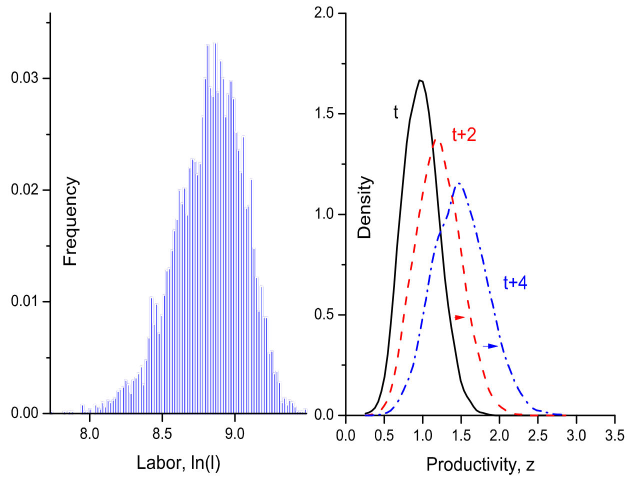


Figure 9: The left panel illustrates the time-invariant firm size distribution. The right panel shows how the productivity distribution evolves over time.

aggregate productivity pulls all firms along as is evident in (2). When a firm's growth in productivity pulls ahead of aggregate growth it loses this slipstream effect, so to speak. Since z' increases in an arithmetic fashion with x and b, growth must decay when $\mathbf{z}$ is held fixed. The distribution for productivity across firms, or Z, is not stationary. Consider a point along a balanced growth path where $\mathbf{z} = 1$. The z's will be distributed on $[0, \overline{z}]$ according to $\widehat{Z}$, where $\overline{z} \equiv (\gamma_d + \gamma_n)/[(\mathbf{g} - 1)]$. Next period $\mathbf{z}$ will have grown to $\mathbf{z}' = \mathbf{gz}$. Now, z will be distributed on the $[0, \mathbf{g}\overline{z}]$ according to $Z' = \widehat{Z}(z/\mathbf{g})$. Note the distribution for the z's is getting stretched rightward over time; i.e., the cumulative distribution function is being defined over an ever increasing domain. In general, if in the current period $Z : [0, z^*] \to [0, 1]$ then for next period $Z' : [0, \mathbf{g}z^*] \to [0, 1]$, where $Z'(z) = Z(z/\mathbf{g})$.

The invariant firm-size distribution for the model is also computed via a Monte Carlo. The unique invariant firm-size distribution associated with the baseline model is shown in the left panel of Figure 9. This distribution resembles a log normal. The coefficient of variation for the distribution is 24.2%. The right panel of Figure 9 illustrates how the productivity distribution shifts rightward over time due to growth in the economy. (Here the histograms calculated from the Monte Carlo simulation are replaced by a fitted density function so the movement could be highlighted). A change in shape of the distribution over time is evident.

Proof of the Existence of a Unique Stationary Firm-Size Distribution. All that needs to be shown is that the stochastic process (44) for $\widehat{z}$ converges weakly to a unique invariant distribution, because e and z are independent of each other and the distribution for e is exogenously given. By Stokey and Lucas (1989), Theorem 12.12, it is sufficient to establish three things. First, the transition operator associated with (44) needs to satisfy the Feller Condition [see Stokey and Lucas (1989), p.

220]. Second, it is required that this transition operator is monotone [Stokey and Lucas (1989), p. 220]. Third, the transition operator must satisfy a mixing condition [Stokey and Lucas (1989), Assumption 12.1].

The stochastic difference equation (44) is continuous in $\widehat{z}$, trivially. It then follows, using Stokey and Lucas (1989) Theorem 8.9 and Exercise 8.10, that a transition operator connected with (44) exists and satisfies the Feller property. Denote this operator by $P(\widehat{z}, B)$, which gives the probability measure connected with a move for z from the point $\widehat{z}$ into the set B. Define the random variable ξ by $\xi \equiv \gamma_d x/\mathbf{g}+\gamma_n b/\mathbf{g}$. Similarly, let $Q(X)$ represent the probability measure connected with drawing a value for $\xi \in X$. To establish monotonicity, consider the integral

$$\int H(\widehat{z}')P(\widehat{z}, d\widehat{z}') = \int H(\frac{1}{\mathbf{g}}\widehat{z} + \xi)Q(d\xi),$$

for any non-decreasing function $H(z')$. Clearly,

$$\int H(\frac{1}{\mathbf{g}}\widehat{z} + \frac{\gamma_d}{\mathbf{g}}x)Q(dx) \geq \int H(\frac{1}{\mathbf{g}}\underline{z} + \xi)Q(d\xi),$$

for all $\widehat{z} > \underline{z}$. Thus, P is monotone.

Finally, turn to the mixing condition. The long-run mean of the above process is $z^* = (\gamma_d E[x] + \gamma_n b)/(\mathbf{g} - 1)$, where $0 < z^* < \overline{z} \equiv (\gamma_d + \gamma_n)/(\mathbf{g} - 1)$. To satisfy the mixing condition, it suffices to show that if the process starts off at $\widehat{z} = 0$ then there exists some chance that it will cross into the interval $[z^*, \overline{z}]$, and analogously if it originates at the $\widehat{z} = \overline{z}$ then there are some odds that it will cross into the set $[0, z^*]$. Clearly, if the process starts off from $\widehat{z} = \overline{z}$ then there are some odds that it will cross into the set $[0, z^*]$. The firm can draw $x = 0$ with strictly positive probability, as can be seen from (43). The same is true for b, too. So, just think about drawing $x = 0$ and $b = 0$ for some prolonged but finite period of time, $T + 1$. Eventually, the process will cross into $[0, z^*]$; this will take a maximum of $T + 1$ periods where $T = \ln(z^*/\overline{z})/\ln(1/\mathbf{g})$. This occurs with probability, $\{(1 - i)[(1 - m_b) + m_b x_a] + i x_k\}^{T+1} \times (1 - p)^{T+1} > 0$. Likewise, if the process starts off from $\widehat{z} = 0$ then there are some odds that it will cross into the set $[z^*, \overline{z}]$. Think about drawing an x shock in the interval $[\underline{x}, 1]$, where $\underline{x} = \max\{E[x] + \varepsilon, x_a, x_k\}$ with $\varepsilon > 0$, together with $b = 1$. This joint event occurs with some strictly positive probability, $\Pr[x \geq \underline{x}] \times p$–again see (43). Imagine drawing this shock for some long, finite period of time. Then, it will take a maximum of $t + 1$ periods for the process to cross into the set $[z^*, \overline{z}]$, where $t = \ln[1 - (z^*/\underline{x})(1 - \rho)/\rho]/\ln\rho$, with $\rho \equiv \gamma_d/\mathbf{g}$. The probability of this occurring is $(\Pr[x > \underline{x}] \times p)^{t+1} > 0$. ■

12 Simplified Model–Supplemental Material

A simplified version of the benchmark model is presented here. The goal is to develop an analytical solution to the simplified model. This solution can be used to connect objects such as matching efficiency in the patent market, η, or the importance of technological propinquity, γ, with the economy's growth rate, $\mathbf{g}$. The development is heuristic in nature and may be useful for readers not familiar with the modeling apparatus employed in the paper.

12.1 Simplifying Assumptions

Some simplifying assumptions are now made in order to solve the model analytically:

- The model is cast in continuous time. This rules out simultaneous events within a period.
- The per-period return for firms, $\Pi(z)$, is given the linear specification $\Pi(z) = \pi z$. Recall that in the benchmark model profits turned out to be linear function in z; cf. (7).
- The law of motion for a firm's productivity is given by $z' = z + \gamma x \mathbf{z}$. Thus, only distance-related patents are allowed; cf. (2).
- The utility function is logarithmic, implying $\varepsilon = 1$.
- The innovation rate, $\mathbf{i} > 0$, is assumed to be exogenous. It is set to unity[25]. The analysis here will focus on the reallocation of ideas (the main point of the research) rather than the creation of ideas.
- All firms can meet a patent agent and buy an idea at any point in time. In the benchmark model only failed inventors could buy from the market. This assumption was made in the benchmark model for technical convenience.[26] The analysis here illustrates that the assumption in the benchmark model is innocuous.
- The seller is assumed to have full bargaining power; $\omega = 1$.
- The propinquity of an idea, x, is drawn from a uniform distribution, rather than the empirical distribution.

[25] In continuous time the arrival rate, i, can potentially be greater than 1, as will be seen.

[26] If one allowed this in the benchmark model, then some firms could acquire two ideas (instead of just one) in a period. The distribution for new ideas would then be over {0,1,2} instead of {0,1}. This can't happen in a continuous time, since the odds of innovating plus buying a patent at a particular instance of time are zero. The results from the continuous time model closely parallel those of the benchmark model.

- There is no expiration date for a patent on the market. In the benchmark model patents on market expired with some probability.

The rest of the model is as before. To make this section self contained, the basic environment is now described in some detail.

12.2 Environment

Time is continuous and household utility is logarithmic. The discount rate is ρ. Along a balanced growth path the real interest rate will be given in standard fashion by $r = \rho + \mathbf{g}$. There is a continuum of firms of measure 1 that are located on a circle, as in the main model. Each firm is defined by its location on that circle and its current productivity z. (Time subscripts are dropped, when possible, to simplify the notation.)

The equilibrium profit of each firm is linear in productivity z such that

$$\Pi(z) = \pi z.$$

A firm's productivity, z, improves through new innovations which arrive at an exogenous rate $\mathbf{i}$. Each new innovation contributes to productivity according to the following law of motion

$$z' = z + \gamma x \mathbf{z}, \tag{45}$$

where $x \equiv 1 - d \in [0, 1]$ is the propinquity of a new innovation to the firm (which is the inverse of a distance) and

$$\mathbf{z} \equiv \int z_j dj$$

is average productivity in the economy. Once an idea arrives it propinquity is drawn from a uniform distribution such that $x \sim U_{[0,1]}$.

The firm has two (mutually exclusive) options:

1. Keep the idea and produce with it.

2. Sell it directly to a patent agent for the price q.

All firms may try to purchase a patent, even those who have successfully innovated.

Denote the number of ideas on the market to be sold by n_a and the number of buyers by n_b. Then, each period, the number of matches is regulated by the matching function

$$M(n_a, n_b) = \eta n_a^{\mu} n_b^{1-\mu}.$$

Since any firm can buy an idea at any time (unlike in the benchmark model) the measure of potential buyers is equal to 1. Therefore, the number of matches at every instant is

$$M(n_a, n_b) = \eta n_a^{\mu}. \tag{46}$$

12.2.1 A Firm's Value Function

The analysis proceeds using the guess and verify technique. Focus on an equilibrium with a cutoff rule for selling a patent; i.e., a firm keeps an idea if $x < x^*$ and sells it otherwise. Conjecture that the cutoff to buy is the same as the one to sell. Discretize time into small intervals of length Δ. Then, the state of a firm can be summarized by the pair $(z, \mathbf{z})$ and the value function for a firm can be written as

$$\begin{aligned} V(z, \mathbf{z}) &= \pi z \Delta + \mathbf{i}\Delta \left[\int_{x^*}^{1} V(z + \gamma x \mathbf{z}_{+\Delta}, \mathbf{z}_{+\Delta})\, dx + \int_{0}^{x^*} [V(z, \mathbf{z}_{+\Delta}) + q]\, dx \right] \\ &\quad + M(n_a, n_b)\, \Delta \int_{x^*}^{1} [V(z + \gamma x \mathbf{z}_{+\Delta}, \mathbf{z}_{+\Delta}) - P(x, z, \mathbf{z})]\, dx \\ &\quad + (1 - r\Delta)[1 - \mathbf{i}\Delta - M(n_a, n_b)\, \Delta]\, V(z, \mathbf{z}_{+\Delta}) + o(\Delta). \end{aligned}$$

Over any time interval Δ the firm collects $\pi z \Delta$ in profit. During the same interval, the firm can receive a new idea with probability $\mathbf{i}\Delta$. (As an aside, this implies that $\mathbf{i}$ can be bigger than one when Δ is sufficiently small.) It will keep the idea if $x > x^*$, and sell the idea to a patent agent for the price q, otherwise. Likewise, a firm can receive an option to buy a patent with the endogenous matching probability $M(n_a, n_b)\, \Delta$. It will decide to buy the idea at price $P(x, z, \mathbf{z})$ if $x > x^*$. This price is determined by a take-it-or-leave-it offer that the patent agent makes. Finally, if neither of these two events happen, which occurs with probability $[1 - \mathbf{i}\Delta - M(n_a, n_b)\, \Delta]$, the firm moves to the next time interval with no change in its own productivity and collects the continuation value $V(z, \mathbf{z}_{+\Delta})$, which is discounted by the multiplicative factor $(1 - r\Delta)$. Note that over the time interval in question that aggregate productivity will evolve from $\mathbf{z}$ to $\mathbf{z}_{+\Delta}$. Finally, observe the presence of some second-order terms denoted by $o(\Delta)$. These terms involve the event that the firm both innovates and buys a patent at the same time, which occurs with probability $\mathbf{i}\Delta M(n_a, n_b)\, \Delta = \mathbf{i} M(n_a, n_b)\, \Delta^2$. As the length of period shrinks these terms will disappear (relative to the other terms in the expression because Δ^2 becomes small relative to Δ).

Divide both sides by Δt. Take the limit as $\Delta \to 0$. Note that $\mathbf{z}_{+\Delta} \to \mathbf{z}$ as $\Delta \to 0$. Now, let

$\dot{V}(z,\mathbf{z}) \equiv \lim_{\Delta\to 0}[V(z,\mathbf{z}_{+\Delta}) - V(z,\mathbf{z})]/\Delta$. Rearranging the resulting expression yields

$$\begin{aligned} rV(z,\mathbf{z}) &= \pi z + \mathbf{i}\left[\int_{x^*}^{1} V(z+\gamma x\mathbf{z},\mathbf{z})\,dx + \int_{0}^{x^*}[V(z,\mathbf{z})+q]\,dx - V(z,\mathbf{z})\right] \\ &\quad + M(n_a,n_b)\int_{x^*}^{1}[V(z+\gamma x\mathbf{z},\mathbf{z}) - V(z,\mathbf{z}) - P(x,z,\mathbf{z})]\,dx + \dot{V}(z,\mathbf{z}). \end{aligned} \tag{47}$$

This continuous-time value function has the following interpretation: The safe return on the left-hand side is equated to the risky return on the right-hand side. Every instant, the firm collects its instantaneous profit, πz. At the rate $\mathbf{i}$ a new idea arrives with some propinquity x. If $x > x^*$, then the firm keeps the idea and its productivity will increase by $\gamma x\mathbf{z}$. Otherwise, the firm sell will sell the idea to the agent at the price q. Finally, $\dot{V}(z,\mathbf{z})$ captures the increase in the firm's market value due to the growth in aggregate productivity, $\mathbf{z}$.

12.2.2 The Price for a Buying Firm

The patent agent makes a take-it-or-leave-it offer to the buyer. Hence, the patent agent will strip all of the surplus from the transaction. The price of patent, $P(x,z,\mathbf{z})$, is then given by

$$P(x,z,\mathbf{z}) = V(z+\gamma x\mathbf{z},\mathbf{z}) - V(z,\mathbf{z}), \text{ when } x \geq x^*. \tag{48}$$

The pricing rule (48) allows the firm's value function (47) to be written as

$$rV(z,\mathbf{z}) = \pi z + \mathbf{i}\left[\int_{x^*}^{1} K(x,\mathbf{z})\,dx + x^* q\right] + \dot{V}(z,\mathbf{z}),$$

where

$$K(x,\mathbf{z}) \equiv V(z+\gamma x\mathbf{z},\mathbf{z}) - V(z,\mathbf{z}) \tag{49}$$

is the value of keeping a new idea that has propinquity x.

12.2.3 Patent Agents and the Price for a Selling Firm

Denote the value function for a patent agent who is about to sell an idea by $A(\mathbf{z})$. This can be expressed as:

$$rA(\mathbf{z}) = \frac{M}{n_a}\int_{x^*}^{1}[P(x,\mathbf{z}) - A(\mathbf{z})]\,dx + \dot{A}(\mathbf{z}), \tag{50}$$

which already incorporates the fact that there will be no trade when $x < x^*$. Also note that the arguments of M are suppressed for clarity. This expression depends on the equilibrium flow of matches per seller, M/n_a, which represents the probability that the agent will find a buyer. Conjecture that the patent agent's value function is linear in aggregate productivity $\mathbf{z}$; i.e.,

$$A(\mathbf{z}) = \mathfrak{a}\mathbf{z}. \tag{51}$$

As in the main text, suppose there is free entry by agents into the market to buy patents from firms. This dictates that the price q will be determined by

$$q = A(\mathbf{z}). \tag{52}$$

The following result holds:

Lemma 3 *(The Value Function for a Firm) Assume that the patent agent's value function, $A(\mathbf{z})$, takes the linear form specified in (51). Then, the firm's value function, $V(z, \mathbf{z})$, is linear in both its own z and the aggregate $\mathbf{z}$ so that*

$$V(z, \mathbf{z}) = \mathfrak{v}_1 z + \mathfrak{v}_2 \mathbf{z}, \tag{53}$$

where

$$\mathfrak{v}_1 = \frac{\pi}{r} \text{ and } \mathfrak{v}_2 = \frac{\mathbf{i}\mathfrak{v}_1\gamma(\frac{1}{2} - \frac{x^{*2}}{2}) + \mathbf{i}x^*\mathfrak{a}}{\rho}.$$

Proof. See Section 3. ■

12.2.4 The Cutoff Rule

The law of motion for the stock of ideas on the market is

$$\dot{n}_a = \mathbf{i}x^* - M(1 - x^*). \tag{54}$$

The first term, $\mathbf{i}x^*$, gives the flow of new ideas into the market, since the rate of innovation is $\mathbf{i}$ and x^* is fraction of new innovations that are put on the market. For the rest of the text, assume that $\mathbf{i} = 1$. The second term, $M[1 - x^*]$, gives the number of ideas that are sold in period. All firms try to buy a patent. The measure of firms is one. They meet an agent with probability M and buy an patent with probability $(1 - x^*)$. The equilibrium number of matches can be found by setting $\dot{n}_a = 0$ in (54) which implies

$$M = \frac{x^*}{1 - x^*}. \tag{55}$$

Clearly the number of matches increases in the cutoff x^*: a higher x^* implies that initial inventors keep less of their ideas since they now have to meet a more stringent threshold. Plugging the number of matches (55) into the matching technology (46) gives the equilibrium stock of ideas on the market or

$$n_a = \left[\frac{x^*}{(1 - x^*)\eta}\right]^{\frac{1}{\mu}}. \tag{56}$$

The number of matches per seller is then

$$m_a \equiv \frac{M}{n_a} = \eta \left[\frac{x^*}{(1-x^*)\,\eta} \right]^{\frac{\mu-1}{\mu}}. \tag{57}$$

The patent agent's value function, $A(\mathbf{z})$, can be solved for now. Using the conjecture (51) that this value function is linear, in conjunction with (50), (55), and (56), generates (the steps are similar to those outlined in the proof of Lemma 3).

$$\mathfrak{a} = \frac{m_a \frac{\mathfrak{v}_1 \gamma}{2} \left(1 - x^{*2}\right)}{\rho + m_a (1 - x^*)}. \tag{58}$$

The only remaining equilibrium variable to be determined is the cutoff, x^*. The marginal seller is indifferent keeping and selling the idea so

$$K(x^*, \mathbf{z}) = q. \tag{59}$$

The value of keeping the idea $K(x^*, \mathbf{z})$ from (49) and (53) is

$$K(x^*, \mathbf{z}) = \frac{\pi}{r} \gamma x^* \mathbf{z}.$$

Now, it has been shown that $A(\mathbf{z}) = \mathfrak{a}\mathbf{z}$, where $\mathfrak{a}$ is expressed in (58), $q = A(\mathbf{z})$, and $\mathfrak{v}_1 = \pi/r$. Using these results in the indifference condition (59) allows the cutoff rule, x^*, to be written as

$$x^* = \frac{m_a \left(1 - x^{*2}\right)}{\rho + m_a (1 - x^*)} \frac{1}{2}, \tag{60}$$

where m_a is expressed in (57).

12.2.5 The Aggregate Growth Rate

The focus can now shift to the aggregate growth rate for the economy, $\mathbf{g}$. Using (45) it is easy to deduce that after a small time interval, Δ, the average productivity will evolve from $\mathbf{z}$ to $\mathbf{z}_{+\Delta}$ according to

$$\mathbf{z}_{+\Delta} = \mathbf{z} + \int_0^1 \left[(\mathbf{i}\Delta + M\Delta) \int_{x^*}^1 x\gamma\mathbf{z}dx \right] dj.$$

To compute the aggregate rate of growth, subtracting $\mathbf{z}$ from both sides, perform the integration on the right-hand side, and divide by $\mathbf{z}\Delta$, which will result in

$$\mathbf{g} = (\mathbf{i} + M)\,\gamma \left(\frac{1}{2} - \frac{x^{*2}}{2} \right).$$

Last, let $\mathbf{i} = 1$ and use equilibrium matching rate given by (55) to write $\mathbf{i} + M = 1 + x/(1-x) = 1/(1-x)$. The above formula can then be rewritten as

$$\mathbf{g} = \quad \gamma \times \left[\underbrace{\frac{1}{2}}_{\substack{\text{non-market} \\ \text{component}}} + \underbrace{\frac{x^*}{2}}_{\substack{\text{market} \\ \text{component}}} \right] \quad . \tag{61}$$

Equation (61) characterizes the role the patent market plays in determining the equilibrium growth rate. Assume there is no patent market so that all firms keep their ideas for themselves. This is equivalent to $x^* = 0$. In this case, since each idea is drawn from a uniform distribution, the average propinquity of an idea to an innovating firm is equal to $E[x] = 1/2$. Moreover, each idea contributes to a firm's productivity by the multiplicative term γ per unit of propinquity. Therefore, the overall growth in productivity will be given by $\gamma/2$.

The role of the patent market comes into play through x^*. Recall that firms keep patents with propinquity $x \geq x^*$ and that patents are drawn from a uniform distance distribution. When the cutoff increases from 0 to x^* the average propinquity of the utilized ideas increases by $x^*/2$ and the average propinquity becomes $(1 + x^*)/2$. Therefore, any market arrangement that increases the threshold x^* contributes positively to economic growth.

12.3 Predictions

In order to achieve further analytical results, assume also that $\mu = 1$. Using (57) and (60) then gives

$$x^* = \frac{\rho}{\eta} + 1 - \sqrt{\left[\frac{\rho}{\eta} + 1\right]^2 - 1}. \tag{62}$$

For use below, note that the cutoff rule, x^*, is decreasing in ρ/η.

Proposition 4 *(Aggregate Growth Rate) The aggregate rate of growth,* $\mathbf{g}$*, is increasing in the productivity gain from a new idea,* γ*, and matching efficiency,* η*, and is decreasing in the rate of time preference,* ρ*.*

The first fact is obvious from (61), while noting from (62) that the cutoff, x^*, is not a function of γ. To see the second fact note that an increase in matching efficiency increases the cutoff, via (62), and therefore growth, though (61). This result is also very intuitive. As the matching technology becomes more efficient, the market value of a patent increases due to more frequent matches. Therefore, the owner becomes more selective when deciding to keep an idea. This increases the cutoff rule x^*.

Hence, markets with more efficient matching technologies grow faster. The last result follows from the fact that more patience translates into a higher expected value from selling the patent as is evident from (58). Hence, firms put more ideas on the market when discount rate decreases and ideas are allocated to better users in this way.

12.4 Proof of Lemma 3

Proof. Substitute the conjecture (53) into (47), while making use of (48), to get

$$r\left[\mathfrak{v}_1 z + \mathfrak{v}_2 \mathbf{z}\right] - \mathfrak{v}_2 \mathbf{z}\mathbf{g} = \pi z + \mathbf{i}\left[\int_{x^*}^{1}\left[\mathfrak{v}_1\left(z + \gamma x \mathbf{z}\right) + \mathfrak{v}_2 \mathbf{z} - \mathfrak{v}_1 z - \mathfrak{v}_2 \mathbf{z}\right] dx\right] + \mathbf{i}x^* q.$$

This expression simplifies to

$$\begin{aligned} r\left[\mathfrak{v}_1 z + \mathfrak{v}_2 \mathbf{z}\right] - \mathfrak{v}_2 \mathbf{z}\mathbf{g} &= \pi z + \mathbf{i}\int_{x^*}^{1} \mathfrak{v}_1 \gamma x \mathbf{z} dx + \mathbf{i}x^* q \\ &= \pi z + \mathbf{i}\mathfrak{v}_1 \gamma \mathbf{z}\left[\frac{1}{2} - \frac{x^{*2}}{2}\right] + \mathbf{i}x^* q. \end{aligned}$$

Equating the terms on the left- and right-hand sides that involve z and $\mathbf{z}$, separately, yields

$$\mathfrak{v}_1 = \frac{\pi}{r}, \text{ and } r\mathfrak{v}_2 \mathbf{z} - \mathfrak{v}_2 \mathbf{z}\mathbf{g} = \mathbf{i}\mathfrak{v}_1 \gamma \mathbf{z}\left[\frac{1}{2} - \frac{x^{*2}}{2}\right] + \mathbf{i}x^* q.$$

Recall that $q = A(\mathbf{z}) = \mathfrak{a}\mathbf{z}$ and that along a balanced growth path $r - \mathbf{g} = \rho$. These facts allow for the formula for $\mathfrak{v}_2$ to be rewritten as

$$\mathfrak{v}_2 = \frac{\mathbf{i}\mathfrak{v}_1 \gamma \left[\frac{1}{2} - \frac{x^{*2}}{2}\right] + \mathbf{i}x^* \mathfrak{a}}{\rho}.$$

■

References

Arora, A., and A. Gambardella (1994): "Evaluating Technological Information and Utilizing It: Scientific Knowledge, Technological Capability and External Linkages in Biotechnology," *Journal of Economic Behavior and Organization*, 24, 91-114.

Arora, A., A. Fosfuri, and A. Gambardella (2001): "Markets for Technology and their Implications for Corporate Strategy," *Industrial and Corporate Change*, 10, 419-451.

Cassiman, B., and R. Veugelers (2006): "In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition," *Management Science*, 52, 68-82.

Chesbrough, H. (2006): "Emerging Secondary Markets for Intellectual Property: US and Japan Comparisons," National Center for Industrial Property Information and Training, Japan.

Elton, J.J., B.R. Shah, and J.N. Voyzey (2002): "Intellectual Property: Partnering for Profit," *McKinsey Quarterly*, 4, 59-67.

Gambardella, A., P. Giuri, and M. Mariani (2006): "Study on Evaluating the Knowledge Economy – What Are Patents Actually Worth?" Directorate-General for Internal Market, European Commission.

Gans, J.S., and S. Stern (2000): "Incumbency and R&D Incentives: Licensing the Gale of Creative Destruction," *Journal of Economics and Management Strategy*, 9, 485–511.

Gans, J.S., and S. Stern (2010): "Is There a Market for Ideas?" *Industrial and Corporate Change*, 19, 805–837.

Hagiu, A., and D. Yoffie (2011): "Intermediaries for the IP Market," Working Paper 12-023, Harvard Business School.

Jovanovic, B. (1993): "The Diversification of Production," *Brookings Papers on Economic Activity, Microeconomics: Microeconomics*, 197-247.

Khan, B.Z. (2013): "Trolls and Other Patent Inventions: Economic History and the Patent Controversy in the Twenty-First Century," Unpublished Manuscript, Bowdoin College.

Khan, Z., and K.L. Sokoloff (2004): "Institutions and Democratic Invention in 19th-Century America: Evidence from 'Great Inventors,' 1790-1930," *American Economic Review*, 94, 395-401.

Lamoreaux, N.R., and K.L. Sokoloff (1997): "Location and Technological Change in the American Glass Industry During the Late Nineteenth and Early Twentieth Centuries," Working Paper 5938, NBER.

Lamoreaux, N.R., and K.L. Sokoloff (1999): "Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries," in *Learning by Doing in Markets, Firms and Countries*, ed. by N.R. Lamoreaux, D.M.G. Raff, and P. Temin. Chicago, IL: University of Chicago Press, 19-57.

Lamoreaux, Naomi R., Kenneth L. Sokoloff, and Dhanoos Sutthiphisal (2009): "The Reorganization of Inventive Activity in the United States during the Early Twentieth Century," Working Paper 15440, NBER.

Lamoreaux, N.R., K.L. Sokoloff, and D. Sutthiphisal (2013): "Patent Alchemy: The Market for Technology in US History," *Business History Review*, 87, 3-38.

Nicholas, T. (2009): "Spatial Diversity in Invention: Evidence from the Early R&D Labs," *Journal of Economic Geography*, 9, 1–31.

Radauer, A, and T. Dudenbostel (2013): "PATLICE Survey - Survey on Patent Licensing Activities by Patenting Firms," Directorate-General for Research and Innovation, European Commission.

Sakkab, N. Y. (2002): "Connect & Develop Complements Research & Develop at P&G," *Research-Technology Management*, 45, 38-45.

Salant, S.W. (1984): "Preemptive Patenting and the Persistence of Monopoly: Comment," *American Economic Review*, 74, 247-250.

Stokey, N.L., and R.E. Lucas, Jr. with E.C. Prescott (1989): *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

Teece, D.J. (1986): "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy," *Research Policy*, 15, 285-305.

Tietze, F. (2010): "A Typology of Technology Market Intermediaries," Working Paper 60, Institute for Technology and Innovation Management, Hamburg University of Technology.Lamoreaux, N.R., and K.L. Sokoloff (1997): "Location and Technological Change in the American Glass Industry During the Late Nineteenth and Early Twentieth Centuries," Working Paper 5938, NBER.

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

Daron Acemoglu† Ufuk Akcigit‡ Murat Alp Celik§

July 30, 2015

Abstract

This paper argues that openness to new, unconventional and disruptive ideas has a first-order impact on creative innovations—innovations that break new ground in terms of knowledge creation. After presenting a motivating model focusing on the choice between incremental and radical innovation, and on how managers of different ages and human capital are sorted across different firms with different degrees of openness to disruption, we provide firm-level, patent-level and cross-country evidence consistent with this pattern. Our measures of creative innovations proxy for innovation quality (average number of citations per patent) and creativity (fraction of superstar innovators, the likelihood of a very high number of citations, and generality of patents). Our main proxy for openness to disruption is the age of the manager—based on the idea that only companies or societies open to such disruption will allow the young to rise up within the hierarchy. Using this proxy at the firm, patent and country level, we present robust evidence that openness to disruption is associated with more creative innovations, but we also show that once the effect of the sorting of young managers to firms that are more open to disruption is factored in, the (causal) impact of manager age on creative innovations is small.

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

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

---

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

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

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

§University of Pennsylvania. Email: celik@sas.upenn.edu.

1 Introduction

This paper investigates the impact of economic and social incentives on "creative innovations," which we identify with the most influential, innovative and original patents. Though there are currently more than half a million patents granted by the US Patent and Trademark Office (USPTO) per year, only a handful are truly transformative in terms of their contribution to society's knowledge and their impact on the organization of production, and probably only a small fraction account for the bulk of the value created.[1] For example, within the field of drugs and medical inventions, there were 223,452 patents between 1975 and 2001, but the median number of citations of these patents within the next five years was four. A few patents receive many more citations, however. One was the patent for "systems and methods for selective electrosurgical treatment of body structures" by the ArthroCare Corporation (with 50 citations), which has also had a major impact on the field by improving many existing surgical procedures and devices used, inter alia, in arthroscopy, neurology, cosmetics, urology, gynecology and laparoscopy/general surgery. Another example is Amazon's patent for "method and system for placing a purchase order via a communications network," which received 263 citations within five years (while the median number of citations within this class is five) and has fundamentally altered online businesses.

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

> "Move fast and break things."

This perspective suggests that societies and organizations that impose a set of rigidly specified rules, discourage initiative and deviations from established norms, shun or even ostracize rebellious behavior, and do not tolerate those that "move fast and break things" will significantly lag behind their more open, "individualistic" or "risk-taking" counterparts in creative innovations—even though they might still be able to function successfully with existing technologies. In the rest of the

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

paper, we thus refer to this constellation of social and economic incentives as *openness to disruption* (short for openness to disruptive innovations, ideas and practices).

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

Our model predicts a reduced-form cross-sectional relationship between manager age and radical innovation. But this relationship does not correspond to the causal effect of manager age on creative innovations. Rather, manager age is both an economically relevant variable and more generally a proxy for openness to disruption, as highlighted by our model where young managers tend to work in firms that are open to radical innovation, but also contribute to the likelihood of radical innovations in such firms. These forces can also be seen from the longitudinal predictions of the model: firms that hire younger managers should subsequently have more creative innovations (because hiring a young manager is associated either with a change in a firm's type or a change in the firm's innovation strategy as it runs out of productive incremental innovation opportunities). But because firms that are more open to disruption need not immediately hire a young manager, the increase in creative innovations can precede the hiring of a younger manager.

The model further clarifies that radical innovations will generate higher quality patents that are more likely to receive a high number of citations and tend to be more general in terms of the range of citations they receive (because they are expanding into new areas), and this provides us with an empirical strategy to measure the creativity of innovations (and present evidence about several aspects of the model's implications).

Our theoretical framework also predicts another relationship we investigate empirically: products with higher sales will encourage even high-type firms and young managers to pursue incremental innovations (because of Arrow's (1962) *replacement effect*), and those with many patents will tilt things in favor of radical innovations (because of diminishing returns and more generally because there is a substantial knowledge base to build upon for such an expansion).

Finally, our model further suggests that institutions or attitudes that ban or discourage expansion into new areas or combinations that have not been previously experimented with can be highly detrimental to radical innovations. Equally, those that prevent young managers from leading companies could slow down creative innovations by failing to use their more recent vintage skills in radical innovations. Such institutions and attitudes typically vary across countries, and this reasoning suggests that similar relationships might be found in the cross-country data.

The bulk of our paper comprises an empirical study of the ideas illustrated by our theoretical model. We investigate whether companies with younger managers engage in more radical and creative innovations. As already noted, manager age is a natural proxy for openness to disruption, since companies with a corporate culture open to disruption are more likely to allow young managers to rise up to the top of the corporate hierarchy.[2]

Our empirical work uses several different measures of creative innovations, all computed from the USPTO data. These are the *average number of citations per patent*; *the fraction of superstar innovators*, which corresponds to the fraction of patents accruing to 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 country or company) that are at the pth percentile of the overall citations distribution (such as the 99th percentile) relative to those that are at the median, thus capturing the likelihood of receiving a very high number of citations normalized by the "median" number of citations; and *generality index*, constructed by Hall, Jaffe and Trajtenberg (2001), which measures the dispersion of the citations that a patent receives from different technology classes. We report several salient and robust pattern using these data.

First, we establish a very robust cross-sectional correlation between CEO (or top management) age and all of our measures of firm-level creative innovation (with or without a variety of firm-level controls). In summary, firms that tend to employ younger CEOs receive a greater number of citations per patent, have a greater fraction of their patents generated by superstar innovators, have more tail innovations, which are at the very high percentiles of the citations distribution, and have more general patents.

Second, we find similar (but somewhat smaller) results when we focus on "within-firm" variation generated by CEO changes: when a younger CEO takes charge, innovations (new patent applications) become more creative. Recall, however, that, as our theoretical analysis highlights, these within-firm results are still a mixture of the sorting effects and the causal effect of manager age on creative innovations.

Third, related to this last point and again consistent with our theoretical model, we show that

[2]Interestingly, in the examples of major innovations mentioned above, these were produced by companies with unusually young leadership. The average age of top managers at ArthroCare Corporation was 41 at the time, and only 33 at Amazon (compared to an average age of 54.84 among Compustat companies).

there is a significant increase in creative innovations *before* a firm switches to a younger CEO, but once it does make the switch, there is a further increase in the creativity of their innovations.

Fourth, we also use the structure of our model, in conjunction with the reduced-form patterns in the data, to shed further light on the relative roles of sorting and the causal effect of manager age on innovation. Namely, we utilize a simple indirect inference procedure to estimate from the reduced-form regression coefficients some of the key parameters of our model, including those governing the causal effect of manager age on creative innovations. This exercise implies that the causal effect of manager age on creative innovations is small and is dwarfed by the sorting effects resulting from the fact that firms that are more open to disruption, and thus more creative, tend to hire younger managers.

Fifth, we exploit the patent-level variation to estimate the separate impacts of CEO and inventor age on the creativity of innovations. Our results indicate that both matter, with roughly similar magnitudes. But we also find that younger CEOs tend to work with younger inventors (though CEO age has a fairly precisely estimated impact even after controlling for inventor age). These two findings, which suggest that firms typically undertake many associated changes while they are switching towards generating more creative innovations, further corroborate our interpretation that much of the cross-sectional (and within-firm) evidence reflects sorting of younger managers to firms with corporate cultures that are more open to disruption.

Finally, we also use the firm-level data to shed light on our model's prediction that firms with greater sales should be less willing to encourage new, potentially disruptive ideas, practices and innovations, while firms that are technologically more advanced, and thus not able to profitably function without engaging in major innovations, should be more likely to encourage this type of disruptive innovation. Our firm-level data enable us to investigate this idea by simultaneously including interactions of CEO age with (log) sales and (log) number of patents of the firm. Though the results here are somewhat less strong than our main findings, they are broadly consistent with the notion that CEO age interacts negatively with sales and positively with the number of patents.

We conclude the paper by showing in Section 5 that the firm-level results aggregate up to the country level, so that countries that employ younger managers appear to have more creative innovations controlling for other factors. For this exercise, we use the average age of (top) managers (e.g., CEO and CFO) in the 25 largest listed companies in the country (when available), which we collected from publicly available sources. We find a fairly stable relationship between manager age and creativity of innovations at the country level as well, suggesting that the forces we emphasize might account for cross-country differences in the type and quality of innovations.

The cross-country context is also useful for us because it provides a corroboration that manager age is indeed capturing practices related to openness to disruption. We do this by utilizing the individualism and uncertainty avoidance indices of "national cultures" constructed by the Dutch social scientist Geert Hofstede.[3] Our results using these indices are similar to those based on average manager age, suggesting that, at least at the country level, our manager age variable is likely to be capturing some aspects related to a society's openness to disruption.

Our paper is related to several literatures. First, we build on and extend the emerging literature on the interplay between micro and macro aspects of innovation. In particular, we build on Klette and Kortum's (2004) model of innovation dynamics by including a choice between radical and incremental innovations, and by incorporating the dimension of matching between managers of different vintages of human capital (age) and type of innovation.[4] The burgeoning empirical literature in this area (e.g., Foster, Haltiwanger and Krizan, 2001, Lentz and Mortensen, 2008, Akcigit and Kerr, 2010, Hurst and Pugsley, 2011, Syverson, 2011, Kogan et al., 2012, Acemoglu et al., 2013) focuses on R&D, patent and productivity dynamics. We depart from this literature both by focusing on the choice between radical (creative) and incremental innovations, and by presenting a detailed analysis of the relationship between creativity of innovations and manager age.

Second, three papers most closely related to our work are MacDonald and Weisbach (2004), Gorodnichenko and Roland (2012) and Fogli and Veldkamp (2013). 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 creative, radical innovations vs. incremental innovations. Gorodnichenko and Roland draw a link between innovation and individualism and provide evidence using Hofstede's individualism data. Despite the similar motivating questions, the approaches of the two papers are very different. While Gorodnichenko and Roland look at aggregate measures of productivity, such as TFP or labor productivity, we focus on creative innovations defined using patent citations data from the USPTO. We therefore first 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. Our main empirical work instead uses the proxy for openness to disruption we have constructed ourselves based on the age of managers across countries and, more centrally, focuses on firm-level and patent-level analysis across US companies. Fogli and Veldkamp also use Hofstede's individualism index in their theoretical and empirical analysis of "individualistic" social

[3]The individualism index is based on Durkheim's (1933) distinction between collectivism and individualism, and measures the extent to which a society functions by relying on loosely knit social ties and thus permits and condones individual actions even when they conflict with collective goals and practices, particularly in a business context. The uncertainty avoidance index, on the other hand, is an inverse proxy for a society's tendency for risk-taking based in part on ideas from Cyert and March's seminal (1963) book.

[4]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).

networks and the diffusion of new technologies, but their emphasis is on how new technologies diffuse over different network structures and their empirical work exploits exposure to different types of diseases to generate cross-country variation in societal network structures.

Third, our work is linked to the small literature on age and creativity. Galenson and Weinberg (1999, 2001), Weinberg and Galenson (2005), Jones and Weinberg (2011) and Jones (2010) provide evidence that a variety of innovators and top scientists are more creative early in their careers, but they also acquire other types of human capital (perhaps generating different types of creativity) later on. Jones (2009) develops a model in which scientists have to spend more time mastering a given area and have to work in teams because the existing stock of knowledge is growing and thus becoming more difficult to absorb and use. Relatedly, Sarada and Tocoian (2013) investigate the impact of the age of the founders of a company on subsequent performance using Brazilian data.[5]

Fourth, our work is related to the literature pioneered by Bertrand and Schoar (2003) and Bloom and Van Reenen (2007, 2010) which investigates the relationship between CEO and manager characteristics and firm performance. Benmelech and Frydman (2014), for example, show that military CEOs pursue more conservative investment and financial strategies (lower investment in R&D), are less likely to be involved in financial fraud, and perform better during times of distress. Bennedsen et al. (2008) show 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. Kaplan et al. (2012) provide evidence from a factor analysis that CEO ability is positively correlated with subsequent firm performance. Also noteworthy in this context is Barker and Mueller (2002), who show that firms with younger CEOs spend more on R&D.

Finally, there is a growing literature on the impact of cultural factors and practices on long-run economic development. The distinction between individualist and collectivist cultures is deep-rooted in sociology (e.g., Durkheim, 1933) and has been widely applied within the sociology, anthropology and psychology literatures (e.g., Parsons, 1949, Kluckhohn and Strodtbeck, 1961, Schwartz, 1994, Triandis, 1995, and Hofstede, 2001). It has been emphasized within the economics literature by Greif (1994), though we are not aware of any other studies emphasizing or empirically investigating the impact of "openness to disruption".[6] More closely related to our focus in this context is Schumpeter's (1934) vision of an innovator as creating disruption, partly in response to economic incentives and partly for psychological motives that lead them to seek challenges and

[5] See also Azoulay, Manso and Zivin (2011) who document the impact of changes in incentives driven by large academic awards and grants on creativity, and Azoulay, Zivin and Wang (2010) who investigate the impact of the death of a very productive co-author on academic productivity.

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

[6] Other aspects of cultural practices have been emphasized as major determinants of economic developments by, among others, Tabellini (2008a,b), Fernandez and Fogli (2009), Guiso, Sapienza and Zingales (2010), and Alesina, Giuliano and Nunn (2011).

deviate from norms, is more closely related to our focus. Traces of this approach can also be seen in Adorno et al.'s (1950) psychological study of authoritarianism, and in McClelland's (1961) and Winslow and Solomon's (1987) approaches to entrepreneurship (see Kirzner, 1997, for a survey). These ideas have been applied in a cross-country context by Shane (1993, 1995), Hofstede (2001), Schwartz (1994), Schwartz and Bilsky (1990) and others. To the best of our knowledge, no other work links these ideas to creative innovations, develops a formal theory along the lines of what we are attempting here, or provides systematic evidence based on firm- or patent-level data.

The rest of the paper is organized as follows. The next section presents our motivating model. Section 3 describes our data sources and variable construction and provides a few basic descriptive statistics. Section 4 presents our main empirical results, which are based on firm-level data. Section 5 returns to the cross-country data and shows that the patterns we identify in the microdata appear to aggregate up to the country level. Section 6 concludes.

2 Motivating Theory

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

2.1 Production

We consider a continuous-time economy in which discounted preferences are defined over a unique final good $Y(t)$. This final good is produced by labor and a continuum of intermediate goods j, each located along a circle, $\mathcal{C}$, of circumference 1. The production technology takes the following constant elasticity of substitution form

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

where $k_j(t)$ denotes the quantity and $q_j(t)$ the quality (productivity) of intermediate good j used in final good production at time t, while L is the total amount of production labor, which is supplied inelastically.

We follow Klette and Kortum (2004) in defining a firm as a collection of leading-edge (best) technologies. A perfectly enforced patent for each leading-edge quality technology is held by a firm, which can produce it at constant marginal cost γ in terms of the unique final good. Because costs and revenues across product lines are independent, a firm will choose price and quantity to maximize profits on each of its product lines. In doing so, it will face an iso-elastic inverse demand derived from equation (1), which can be written, suppressing time arguments, as:

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

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

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

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

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

Equilibrium profits for a product line with technology q_j are

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

where the second line defines π.

2.2 Managers

In addition to workers, the economy is also populated by managers, who play both 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 age a, or equivalently by her birth date b. Denoting the death rate of managers by δ, the fact that the measure of managers is constant at M implies that the age distribution of managers is simply given by an exponential distribution, i.e., the fraction of managers who are below the age a is $1-e^{-\delta a}$.[7]

When a manager is born, she acquires the knowledge associated with the average technology in the period in which she is born, giving her a knowledge base of

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

Similarly, we denote the current period's knowledge stock—current average technology—by

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

Managers will be hired by monopolists to manage production and innovation in their leading-edge products. In equilibrium, managers will be paid a wage $w_{b,t}$ as a function of the current period's technology, $\bar{q}_t$, and their knowledge, $\bar{q}_b$. We assume that $M<1$, which implies that the measure of managers is less than the measure of product lines in the economy, so some product lines will

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

not use a manager. This simplifies the analysis by providing a simple boundary condition for the determination of equilibrium wages of managers. We also assume that M is not too small, which will ensure that all firms that need a manager for a "radical innovation," as described next, are able to hire one (one can take $M \to 1$ without any loss of generality).

2.3 Corporate Culture and Innovation Dynamics

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

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

Incremental Innovation Incremental innovations improve the productivity of a product line within the current *technology cluster*.[9] 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 will run into diminishing returns. We model this by assuming that the additional productivity improvements generated by an innovation decline in the number of prior incremental innovations within a technology cluster. In addition, again for the same reason, incremental innovations build on a narrow technology base and create improvements only over this base. This implies that, as illustrated in Example 1 below, incremental innovations will have few citations and limited "generality" (captured by the dispersion of citations they receive from different technology classes as we discuss further below).

We assume that all firms (regardless of their type) can successfully innovate incrementally at the exogenous rate $\xi > 0$. The n^{th} incremental innovation in a technology cluster improves the current productivity of product line j by a step size $\eta_n(q_j, \bar{q}_t)$, where q_j is the current productivity

[8] We assume that there are no switches from high type to low type to simplify the expressions and the analysis.
[9] Our modeling of technology clusters follows Akcigit and Kerr (2010) and Abrams et al (2013).

of the technology, and $\bar{q}_t$ is the current period's technology, and

$$\eta_n(q_j, \bar{q}_t) = [\kappa \bar{q}_t + (1-\kappa)\, q_j]\, \eta \alpha^n \tag{3}$$

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 decline geometrically, at the rate α, in the number of prior incremental innovations in the technology cluster.

Denoting by t_n the time of the n^{th} incremental improvement for product line j, the evolution of the technology of product line j in a technology cluster that started with productivity q_j^0 after n incremental innovations can then be written as

$$\begin{aligned} q_j^n &= q_j^0 + \sum\nolimits_{i=0}^{n} \left[\kappa \bar{q}\,(t_i) + (1-\kappa)\, q_j^0\right] \eta \alpha^i \\ &= q_j^0 \left[1 + (1-\kappa)\, \eta \frac{1-\alpha^{n+1}}{1-\alpha}\right] + \eta\kappa \sum\nolimits_{i=0}^{n} \alpha^i \bar{q}\,(t_i)\,. \end{aligned}$$

Radical Innovations Radical innovations 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. Because they create new technology clusters, radical innovations tend to receive more citations, are more likely to have a very high number of ("tail") citations, and have greater generality.

If there is a radical innovation in a particular product line, the innovator will initiate a new technology cluster in a different product line (and will still keep its original product line). The creation of a new technology cluster generates a larger improvement on current technology, and also 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. In particular, 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

[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 would not affect any of our results.

technology (now under the control of the innovating firm and manager), with productivity

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

where the superscript 0 denotes 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 radical innovations for firms attempting such radical innovations, as we describe next.

A firm of type θ has a baseline flow rate of radical innovation (regardless of whether they are 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 also have a flow rate of radical innovation equal to

$$\Lambda\theta\bar{q}^a, \tag{4}$$

where

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

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 implies 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.

Moreover, since both high- and low-type firms have the same rate of success, at the rate ξ, when they attempt incremental innovations, our model also implies that θ captures the *comparative advantage* of firms for radical innovation. In addition, young managers also have a *comparative advantage* in radical innovation—since the contribution of the manager of age a to cost reductions is the same for all firms, 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 permit only the implementation of certain radical innovations, thus making successful innovations less likely.[12]

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

[12] In particular, in the context of our modeling of product lines along the circle $\mathcal{C}$, we may assume that such sanctions

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

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

$$\text{First product line:} \quad \Big| \; \underbrace{\overbrace{P_1}^{\eta_0}\;\overbrace{P_2}^{\eta_1}\;\overbrace{P_6}^{\eta_2}\;\overbrace{P_7}^{\eta_3}}_{\textit{Tech Cluster 1}} \; \Big| \; \underbrace{\overbrace{P_{11}}^{\eta_0}\;\overbrace{P_{12}}^{\eta_1}}_{\textit{Tech Cluster 2}} \; \Big| \; \underbrace{\overbrace{P_{13}}^{\eta_0}\;\overbrace{P_{14}}^{\eta_1}\;\overbrace{P_{15}}^{\eta_2}}_{\textit{Tech Cluster 3}}$$

$$\text{Second product line:} \quad \Big| \; \underbrace{\overbrace{P_3}^{\eta_0}\;\overbrace{P_4}^{\eta_1}\;\overbrace{P_5}^{\eta_2}}_{\textit{Tech Cluster 1}} \; \Big| \; \underbrace{\overbrace{P_8}^{\eta_0}\;\overbrace{P_9}^{\eta_1}\;\overbrace{P_{10}}^{\eta_2}}_{\textit{Tech Cluster 2}}$$

In this example, P_n denotes the n^{th} patent registered at the patent office and η_n denotes the step size as described in equation (3). The first technology cluster starts with a radical innovation associated with a patent P_1. The productivity improvement due to this patent is η_0. Subsequently a new incremental innovation in this technology cluster, with patent P_2, follows on P_1, increasing productivity by another $\eta_1 < \eta_0$. After this innovation, there is a radical innovation P_3 in the second product line, followed by two subsequent incremental innovations P_4 and P_5. Since P_5 and P_6 are second incremental innovations in their technology clusters, they increase productivity by $\eta_2 < \eta_1$. Note that P_1, P_3, P_8, P_{11} and P_{13} are radical innovations starting new technology clusters. As described above, these come from innovations in other product lines operated by high-type firms. Suppose also that the firm operating technology cluster 1 with patent P_7 is a high-type firm, and successfully undertakes a radical innovation after P_7, launching a new technology cluster on a different product line (shown above as patent P_8*).*

Consider next the patterns of citation resulting from these innovations. 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 table. (Alternatively, such patents might also cite patterns from previous technology clusters on the same product line, with very similar patterns). In addition, it is also plausible that, because a radical innovation is recombining ideas from its own product line and the product line on which it is building, it should be citing the fundamental ideas encapsulated in the patents that initiated the two technology clusters. For this reason, patents P_8, P_{11}, and P_{13} cite the patents initiating the previous technology cluster in this product line as well as the patent

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 correspond to $\Lambda = 1/2$, 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.

initiating the most recent technology cluster in their own product line. The next table shows this citation pattern for the first five patents.

Cited	***Citing***
$P_1:$	$P_2, P_6, P_7, P_8, P_{11}$
$P_2:$	P_6, P_7
$P_3:$	P_4, P_5, P_8
$P_4:$	P_5
$P_5:$	*none*

For example, P_2 builds only on P_1 and thus only cites P_1, and is in turn cited by P_6 and P_7. P_1 is cited not only by the patents that build on itself within the same product line, P_2, P_6, P_7 and P_{11}, but also by P_8 because this new patent comes out of recombining ideas based on this technology cluster and those in some other product line. This pattern then implies that radical innovations will receive more citations and will receive more "general" citations as well. They will also be heavily overrepresented among "tail innovations," meaning among patents receiving the highest number of citations. These are the patterns we will explore in our empirical work.

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

2.4 Value Functions and Firm Maximization

Recall that a firm makes the innovation decision in each of its product lines to maximize its present discounted value, which we denote by $W_s\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)$ where $s \in \{H, L\}$, $\overrightarrow{q_f}$ is the vector of productivities of the firm, $\overrightarrow{n_f}$ is the vector of the number of incremental innovations in each of these product lines, i.e., $\overrightarrow{q_f} \equiv \left\{q_{f,j_1}, q_{f,j_2}, ..., q_{f,j_{m_f}}\right\}$, and $\overrightarrow{n_f} \equiv \left\{n_{f,j_1}, n_{f,j_2}, ..., n_{f,j_{m_f}}\right\}$, and m_f denotes the number of product lines that firm f is operating.[13] The value function for a low-type firm can be written as

$$rW_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) - \dot{W}_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) = \sum_{m=1}^{m_f} \left[\begin{array}{c} \max_a \left\{\pi q_{f,j_m} + \bar{q}_t f(a) - w_{a,t}\right\} \\ +\xi \left[\begin{array}{c} W_L \left(\begin{array}{c} \overrightarrow{q_f} \backslash \left\{q_{f,j_m}\right\} \cup \left\{q_{f,j_m} + \eta_{n_{f,j_m}+1}\right\}, \\ \overrightarrow{n_f} \backslash \left\{n_{f,j_m}\right\} \cup \left\{n_{f,j_m} + 1\right\} \end{array} \right) \\ -W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) \end{array} \right] \\ +\tau\left[W_L\left(\overrightarrow{q_f} \backslash \left\{q_{f,j_m}\right\}, \overrightarrow{n_f} \backslash \left\{n_{f,j_m}\right\}\right) - W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)\right] \end{array} \right] \tag{5}$$
$$+\varphi\left[W_H\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right) - W_L\left(\overrightarrow{q_f}, \overrightarrow{n_f}\right)\right].$$

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

We can explain the right-hand side of this value function as follows: for each product line $m = 1, ..., m_f$, the firm receives a revenue stream of $\pi q_{f,j_m}$ as a function of its productivity in this product line, q_{f,j_m}. In addition, it has a choice of the age of the manager it will hire to operate this product line (formally choosing $a \in \mathbb{R}_+ \cup \{\emptyset\}$, which is suppressed to save on notation), and if the manager's age is a, it will have additional revenue/cost savings of $\bar{q}_t f(a)$ and pay the market price for such a manager of age a at time t, $w_{a,t}$. Summing over all of its product lines gives the current revenues of the firm. In addition, the firm can undertake an innovation on the basis of each of its active product lines. Since we are looking at a low-type firm, all innovations will be incremental, thus arriving at the rate ξ. When such an innovation happens in product line m that has already undergone n_{f,j_m} incremental innovations, the mth element of $\overrightarrow{q_f}$ changes from q_{f,j_m} to $q_{f,j_m} + \eta_{n_{f,j_m}+1}$ and n goes up by one. We represent this with the arguments of the value function changing to $\overrightarrow{q_f} \backslash \{q_{f,j_m}\} \cup \left\{q_{f,j_m} + \eta_{n_{f,j_m}+1}\right\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\} \cup \{n_{f,j_m} + 1\}$ (and the firm relinquishes its current value function $W_L(\overrightarrow{q_f}, \overrightarrow{n_f})$). The firm might also lose one of its currently active product lines to creative destruction, which happens at the endogenous rate τ (which will be determined in Section 2.6), and in that case, the firm's value function changes from $W_L(\overrightarrow{q_f}, \overrightarrow{n_f})$ to $W_L(\overrightarrow{q_f} \backslash \{q_{f,j_m}\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\})$ (i.e., $\overrightarrow{q_f}$ changes $\overrightarrow{q_f} \backslash \{q_{f,j_m}\}$ and $\overrightarrow{n_f}$ to $\overrightarrow{n_f} \backslash \{n_{f,j_m}\}$). Finally, the last term is due to the fact that a low-type firm switches to high-type at the flow rate φ, in which case it relinquishes its current value function and begets the value function of a high-type firm, $W_H(\overrightarrow{q_f}, \overrightarrow{n_f})$.

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

$$
\begin{aligned}
& rW_H(\overrightarrow{q_f}, \overrightarrow{n_f}) - \dot{W}_H(\overrightarrow{q_f}, \overrightarrow{n_f}) \qquad (6) \\
&= \sum_{m=1}^{m_f} \max \left\{ \begin{array}{c} \pi q_{f,j_m} \\ + \max_a \left\{ \bar{q}_t f(a) - w_{a,t} + \xi \left[\begin{array}{c} W_H \left(\begin{array}{c} \overrightarrow{q_f} \backslash \{q_{f,j_m}\} \cup \left\{ q_{f,j_m} + \eta_{n_{f,j_m}+1} \right\}, \\ \overrightarrow{n_f} \backslash \{n_{f,j_m}\} \cup \{n_{f,j_m} + 1\} \end{array} \right) \\ -W_H(\overrightarrow{q_f}, \overrightarrow{n_f}) \end{array} \right] \right\}; \\ \pi q_m + \max_a \left\{ \bar{q}_t f(a) + \Lambda \theta_H \bar{q}^a \left[\begin{array}{c} \mathbb{E} W_H\left(\overrightarrow{q}_f \cup \{q_{j'} + \eta_0\}, \overrightarrow{n}_f \cup \{0\} \right) \\ -W_H(\overrightarrow{q}_f, \overrightarrow{n}_f) \end{array} \right] - w_{a,t} \right\} \end{array} \right\} \\
&+ \sum_{m=1}^{m_f} \tau \left[W_H(\overrightarrow{q_f} \backslash \{q_{f,j_m}\}, \overrightarrow{n_f} \backslash \{n_{f,j_m}\}) - W_H(\overrightarrow{q_f}, \overrightarrow{n_f}) \right] \\
&+ \sum_{m=1}^{m_f} \psi \Lambda \theta_H \left[\mathbb{E} W_H \left(\begin{array}{c} \overrightarrow{q}_f \cup \{q_{j'} + \eta_0\}, \\ \overrightarrow{n}_f \cup \{0\} \end{array} \right) - W_H(\overrightarrow{q}_f, \overrightarrow{n}_f) \right]
\end{aligned}
$$

The intuition for this value function is very similar to (5) 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 (5). 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 (4). In the case of a successful radical innovation, the value of the firm changes to $\mathbb{E}W_H\left(\overrightarrow{q_f} \cup \left\{q_{j'} + \eta_0\right\}, \overrightarrow{n_f} \cup \{0\}\right)$, where the expectation is over a product line drawn uniformly at random upon which the radical innovation will build.

The next proposition shows that, as in Klette and Kortum (2004) and Acemoglu et al. (2013), these value functions can be decomposed into sums of value functions defined at the product-line level.

Proposition 1 *The value functions in* (5) *and* (6) *can be written as*

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

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 $s \in \{H, L\}$ such that

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

and

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

where $\mathbb{E}V_H(\bar{q}_t)$ denotes the expected value of a radical innovation when the aggregate technology level is $\bar{q}_t$.

Proof. Both of these value functions can be derived straightforwardly by conjecturing the above forms and verifying the conjecture. ■

2.5 Stationary Equilibrium With $\kappa = 1$

We now characterize the stationary equilibrium of this economy in the case where $\kappa = 1$—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.

Value Functions in 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. Let us first consider the value of a product line for a low-type firm. From Proposition 1, we can focus on the decisions and the value function of such a firm at the product line level, and the relevant value function is given by (7).

Since some firms will not hire managers (as $M < 1$), all firms not undertaking radical innovations must be indifferent between hiring and not hiring a manager, which implies that the equilibrium wage for managers, employed by firms engaged in incremental innovations, satisfies the boundary condition:

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

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

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

Solving this value function gives an explicit characterization of the value function of low-type firms as shown in the next proposition.

Proposition 2 *Let us assume that the value function for a high-type firm takes the following form:* $V_H(q_j, n) = Aq_j + \tilde{B}(n)\bar{q}_t$. *Then the value function of a product line operated by a low-type firm,* (7) *takes the following form*

$$V_L(q_j, n) = Aq_j + B(n)\bar{q}_t \tag{10}$$

where

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

and $\tilde{B}(n)$ *is defined in Proposition 3 below.*

Proof. See the Appendix. ■

The form of the value function in (10) is intuitive. It depends linearly on current productivity, q_j, since this 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) since a higher n implies that the next incremental innovation will increase productivity by less—and incremental innovation is the only type of innovation that a low-type firm can undertake.

We next turn to the value of a product line operated by a high-type firm, which differs from (7) because high-type firms have to decide whether to engage in incremental or radical innovation, given by (8) above. Because (4) implies that younger managers have comparative advantage in radical innovation, it follows straightforwardly that there will exist a maximum age a^* such that only managers below this age will work in firms attempting radical innovation. Moreover, the maximization over the age of the manager in (8) implies that such a firm must be indifferent between hiring any manager younger than a^*. This implies:

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

Note that the oldest manager hired for radical innovation earns (from expression (9))

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

Hence

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

This wage schedule highlights that, in general, 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]

Now substituting for (11) in (8), we obtain a simplified form of the value function of a product line operated by a high-type firm as

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

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

Proposition 3 *The value function in* (8) *takes the following form*

$$V_H(q_j,n) = Aq_j + \bar{q}_t\tilde{B}(n), \tag{12}$$

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

where A and $B(n)$ are as defined in Proposition 2) and $\tilde{B}(n)$ is given by

$$[r-g+\tau]\,\tilde{B}(n) \;=\; \psi\left[A(1+\eta)+\tilde{B}(0)\right] + \begin{cases} \xi\left[\tilde{A}\eta\alpha^{n+1}+\tilde{B}(n+1)-\tilde{B}(n)\right] & \text{for } n<n^* \\ \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)\tilde{A}+\tilde{B}(0)\right] & \text{for } n\geq n^* \end{cases}, \tag{13}$$

where $n^*\in\mathbb{Z}_{++}$ is the number of incremental innovations within a technology cluster at which there is a switch to radical innovation given by

$$n^*=\lceil n'\rceil \;\text{ such that }\; \xi\left[A\eta\alpha^{n'+1}+\tilde{B}(n'+1)-\tilde{B}(n')\right]=\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]. \tag{14}$$

Proof. See the Appendix. ■

The intuition for this high-type value function is similar to that for Proposition 2, 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 (and from that point on n will no longer be relevant). This critical value n^* is given by (14), which designates the smallest integer after n' where n' equates the value of attempting an additional incremental innovation to the value of attempting a radical innovation (the notation $\lceil n\rceil$ denotes the next integer after n).

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

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

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

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

$$w_{a,t}=\begin{cases} f(a)\,\bar{q}_t & \text{for} \quad a>a^* \\ [f(a)+\Lambda\theta_H(\bar{q}^a-\bar{q}^{a^*})v]\bar{q}_t & \text{for} \quad a\leq a^* \end{cases}. \tag{15}$$

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

Equilibrium Characterization The next proposition provides the characterization of the stationary equilibrium.

Proposition 4 *Low-type firms (those with $\theta = \theta_L$) always hire "old" managers (those with $a > a^*$ or $b < b_t^*$), pursue incremental innovations and never generate radical innovations.*

High-type firms (those with $\theta = \theta_H$) pursue incremental innovations on product lines less than n^* prior incremental innovations, where n^* is given by (14), and hire "old" managers (those with $a > a^*$ or $b < b_t^*$). They pursue radical innovations on product lines with more than n^* and 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).

Proof. This result directly follows from Propositions 2 and 3. ■

Empirical Implications Our empirical work is inspired by Proposition 4. As explained above, radical innovations will be associated with greater indices of our measures of creative innovations (innovation quality, tail innovations, superstar fraction, and generality). We will first investigate the cross-sectional relationship between manager (CEO) age and creative innovations. In these cross-sectional regressions, manager age is taken to be a proxy of a corporate culture that is more open to disruption. Therefore, from Proposition 4, we expect a negative cross-sectional relationship between manager age and creative innovations. As just stressed, this cross-sectional relationship does not correspond to the "causal effect" of manager age on creativity of innovations (which would apply if we varied manager age holding the firm's corporate culture constant); in particular, it also reflects the sorting of younger managers to corporate cultures that are open to disruption (and thus more conducive to creative innovations).

Our model also has longitudinal implications—that is, implications about how manager age and creativity of innovations vary over time at the firm level—which 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 4.

Recall that low-type firms always engage in incremental innovations and never generate radical innovations. In contrast, high-type firms may attempt a radical innovation depending on how many prior incremental innovations they have had on a product line.

- For 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, however, 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. In this case, the average rate of radical innovation across product lines with $n \geq n^*$ and operated by high-type firms can be computed using the aforementioned fact that the age distribution of managers is given by the exponential distribution, as

$$\psi\Lambda\theta_H + \frac{1}{F(a^*)}\int_0^{a^*} \Lambda\theta_H \bar{q}^a dF(a) = \psi\Lambda\theta_H + \frac{\Lambda\theta_H\delta}{g+\delta}\frac{\left[1-e^{-(g+\delta)a^*}\right]}{\left[1-e^{-\delta a^*}\right]}. \tag{16}$$

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, keeping its old manager.[15] In the process, it will generate radical/creative innovations at the flow rate $\psi\Lambda\theta_H$ as noted above. When it reaches $n = n^*$, it will hire a young manager, switch to a radical innovation strategy, and at that point, its rate of radical/creative innovations will increase, on average, from $\psi\Lambda\theta_H$ to the expression in (16). In contrast, if the product line of the firm at the time of switching to high-type has had $n \geq n^*$ incremental innovations, it will immediately hire a young manager, pursue a radical innovation strategy, and have radical innovations at the flow rate (16).

This discussion implies that when we focus on the relationship between within-firm changes in manager age and creative innovations, we expect to find two regularities. First, when a firm switches from an older to a younger manager, this should be associated with an increase in creative innovations. Second, firms that switch from an older to a younger manager should, on average, experience an increase in creative innovations even *before* the switch. Namely, before the actual switch to a younger manager, the increase in creative innovations approximately corresponds to $\psi\Lambda\theta_H$, whereas following the switch to a younger manager, there will be a further increase in creative innovations corresponding to the second term in (16). Note, however, that even this further increase following the switch to a younger manager does not correspond to the causal effect of manager age on creative innovations for several reasons; first, for firms with $n \geq n^*$, both events will be taking place at the same time; second, even for firms with $n < n^*$, the impact following the switch to a younger manager still contains the sorting effect and also depends on the matching patterns between managers and firms as indicated by the presence of the terms representing the age distribution of managers; and third, strictly speaking to obtain the causal effect, we need to keep the firm type and number of prior incremental innovations constant, and just change (exogenously) manager age—and this is the exercise will will perform in Section 4.4 below.

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

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

society discouraging radical innovations.

2.6 General Equilibrium and the Stationary Distribution of Products

We next characterize the stationary distribution of product lines in this economy in terms of the types of their owners and also in terms of the prior number of incremental innovations, and then use these distributions to determine the aggregate growth rate of the economy in the stationary equilibrium.

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

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

Consider the first line corresponding to $n = 0$. Outflows from this state, products with $n = 0$ operated by high-type firms, come from two sources. First, there is creative destruction coming from low-type firms, which takes place at the rate τ^L per product line (and hence multiplied by μ_0^H). Second, the high-type firm operating this product line has a successful incremental innovation, which takes place at the rate ξ (similarly multiplied by μ_0^H). Inflows into this state are due to creative destruction coming from the high-type firm, which takes place at the rate τ^H (multiplied by the fraction of all product lines except those that are already in this state, thus $\left(1-\mu_0^H\right)$), or due to a low-type firm operating a product line at $n = 0$ changing its type to high, which adds the flow rate $\varphi\mu_0^L$. The other lines are explained similarly, except that creative destruction coming from high-type firms also generates outflows for $n > 1$, and there is no inflow coming from incremental innovations for product lines with $n > n^*$ since high-type firms switch to radical innovation at $n = n^*$.[16]

The flow equations for the low-type product lines can be written similarly and have a similar intuition

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

[16] These equations are written under the assumption that $n^* > 0$. When $n^* = 0$, high-type firms never undertake incremental innovations, and thus the flow equations become $\tau^L\mu_0^H = (1-\mu_0^H)\tau^H + \varphi\mu_0^L$ for $n = n^* = 0$, and $\left(\tau^L + \tau^H\right)\mu_n^H = \varphi\mu_n^L$ for $n > n^* = 0$.

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

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

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

$$\tau = \tau^L + \tau^H.$$

To derive the aggregate growth rate, we combine (1) with (2) to obtain

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

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

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

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

$$g = \eta\tau + \eta\xi\left[\sum\nolimits_{n=0}^{n^*-1} \mu_n^H \alpha^{n+1} + \sum\nolimits_{n=0}^{\infty} \mu_n^L \alpha^{n+1}\right]. \tag{17}$$

2.7 Equilibrium With $\kappa < 1$

In this subsection, we turn to the general case with $\kappa < 1$. We will show that the structure of the equilibrium is similar, 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.

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

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

and

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

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

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

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

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

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

and

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

Proposition 5 *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.*

Proof. See the Appendix. ■

This proposition thus establishes that in this generalized setup (with $\kappa < 1$), the main results from Proposition 4 continue to hold, but in addition we obtain the 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. We will investigate this additional implication in our firm-level analysis.[17]

[17] This result is related to the idea of "disruptive innovations" proposed in Christensen's *The Innovator's Dilemma* (1997). This result also clarifies that our potential answer to the innovator's dilemma, consistent both with Arrow's 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 new ideas, practices and innovations.

3 Data and Variable Construction

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

3.1 Data Sources

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

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

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

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

Cross-Country Data on Manager Age We also collected data on the age of the CEOs and CFOs of the 25 largest listed companies for 37 countries. We selected the top 25 companies,

[18] Details on the assignment procedure are provided at https://sites.google.com/site/patentdataproject/.
[19] We drop observations where reported CEO age is less than 26.

when available, according to the Financial Times' FT-500 list, which ranks firms according to their market capitalization. We completed the list by using information from transnationale.org when the FT-500 did not include 25 companies for a country. We then obtained the age of the CEOs and CFOs from the websites of the companies. Overall, our dataset has on average 20 companies and 31.6 managers (CEO or CFO) per country.

National Culture Dimensions The Dutch social scientist Geert Hofstede devised five different indices of national culture: power distance, individualism vs. collectivism, masculinity vs. femininity, uncertainty avoidance, and long-term orientation. The initial survey was conducted among IBM employees in 30 countries to understand cross-country differences in corporate culture. This work has been expanded with additional surveys that have been answered by members of other professions and expanded to 80 countries (see Hofstede, 2001, and Hofstede et al., 2010).[20] We use the individualism and uncertainty avoidance measures below.

The individualism measure is defined as "a preference for a loosely-knit social framework in which individuals are expected to take care of themselves and their immediate families only." A low individualism score is indicative of a more collectivist society, where social safety networks are more common and individuals are influenced by collective goals and constraints.

The uncertainty avoidance measure expresses the degree to which the members of a society seek to avoid uncertainty and ambiguity. Countries with a higher score are more rigid in terms of belief and behavior and are more intolerant of unorthodox ideas. On the other end of the spectrum, societies with a low score are more welcoming to new ideas and value practice above principles. Both the individualism and the uncertainty avoidance indices are normalized to lie between 0 and 1.

Other Data Sources We use the average years of schooling in secondary education from the Barro-Lee dataset as a proxy of the human capital of a country.[21] We also use real GDP per capita numbers and R&D intensity from the World Bank's World Development Indicators database.

In our baseline analysis, we focus on a *balanced panel* of firms, with complete information on all variables used in our cross-sectional analysis. To maximize the number of observations in this balanced panel, we focus on the years between 1995 and 2000, so citation and patents information in our baseline results come from 1995-2000 (with patents classified according to their year of *application*). We then extend our analysis to an *unbalanced panel* spanning 1992-2004 (we cannot go earlier than 1992 because our manager age data start at the state, and we cannot go further than

[20] http://geert-hofstede.com/national-culture.html
[21] http://www.barrolee.com/data/dataexp.htm. See Barro and Lee (2013) for details.

2004 as we need a subsequent window during which to measure citations). We also use citations from 1995-2000 in our cross-country analysis.

3.2 Variable Construction

Innovation Quality Our baseline measure of innovation quality is the number of citations a patent received as of 2006. We also use the truncation correction weights devised by Hall, Jaffe, and Trajtenberg (2001) to correct for systematic citation differences across different technology classes and also for the fact that earlier patents will have more years during which they can receive citations (we also experiment with counting citations during a five-year window for each patent). Based on this variable, an average innovation quality variable is generated at the *company* $\times$ *year* and *country* $\times$ *year* levels. For our cross-country dataset, the country of the assignee is used to determine the country to which the patent belongs.

Superstar Fraction A superstar inventor is defined as an inventor who surpasses his or her peers in the quality of patents generated as observed in the sample. A score for each unique inventor is generated by calculating the average quality of all the patents in which the inventor took part. All inventors are ranked according to this score, and the top 5% are considered to be superstar inventors. The superstar fraction of a company or country in a given 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 of the patent). The country of the inventor is determined according to the country of the patent assignee.

Tail Innovations The tail innovation index is defined as the fraction of patents of a firm or country that receive more than a certain number of citations (once again using the truncation correction weights of Hall, Jaffe and Trajtenberg, 2001). Namely, let $s_{ft}(p)$ denote the number of the patents of a firm (or country) that are above the p^{th} percentile of the year t distribution according to citations. Then, the tail innovation index is defined as

$$\text{Tail}_{ft}(p) = \frac{s_{ft}(p)}{s_{ft}(0.50)},$$

where $p > 0.50$. This is of course also equivalent to the ratio of the number of patents by firm f at time t with citations above the p^{th} percentile divided by the number of patents by firm f at time t with citations above the median (and is not defined for firms that have no patents with citations above the median). For our baseline measure of tail innovations, we choose $p = 0.99$, so that our measure is the fraction of patents of a firm or country that are at the 99th percentile of citations divided by the fraction of patents that are at the median of citations. The reason we include $s_{ft}(0.50)$ in the denominator is that we would like to capture whether, controlling for "average" innovation output, some companies, innovators or countries have the tendency for generating "tail innovations" with very high citations.[22]

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 it gives to other patents. Both indices are averaged across all of the patents of a firm or a country to obtain our firm-level and cross-country generality and originality indices. The patent classes used are the 80 two-digit International Patent Classification (IPC) classes.

3.3 Descriptive Statistics

Panel A of Table 1 provides descriptive statistics for our balanced firm, unbalanced firm and cross-country samples. Since we focus on regressions weighted by the number of patents held by a company or country, all statistics are weighted by the number of patents. We multiply our indices for tail innovation, superstar fraction, generality, and R&D intensity by 100.

The table shows that average manager age is 52.3 in our firm-level (balanced or unbalanced Compustat) sample and 56.1 in our cross-country sample, while average CEO age is 55.3 in the balanced sample and 55.5 in the unbalanced sample. The comparison of our average number of citations per patent, superstar fraction, tail innovation, and generality indices shows that, as expected, our Compustat firms have higher values than the average country.

Panels B and C present the firm-level and cross-section country correlations between our main measures of creativity of innovations, which are quite highly correlated except for the generality index at the firm level. Panel D of Table 1 presents the correlation between our three cross-country indices of openness to disruption. These three indices are also highly correlated.

[22] If we do not include the correction in the denominator, all of the results reported below continue to hold, and are in fact stronger. When we turn to patent-level regressions, we will not be able to include such a correction (since there are no other patents that can be included in the denominator).

4 Firm-Level and Patent-Level Results

Our main empirical results exploit firm-level variation in manager age across Compustat companies. Recall that in our theory manager age is in part an indicator of a corporate culture that is open to disruption (because high-type firms that have a competitive advantage in radical innovation select to hire younger managers). But there is also a causal effect of manager age on creative innovations since, conditional on being employed by such a firm, a young manager contributes to radical/creative innovations (because of her more recent knowledge stock). Motivated by this reasoning, in this section we start with the cross-sectional relationship between firm-level measures of creative innovations and manager age (emphasizing throughout that our estimates do not necessarily correspond to the causal effect of manager age on creative innovations).[23] We then turn to a more direct investigation of the effect of manager age on creative innovations, focusing on regressions that exploit "within-firm" variation, and also investigate the timing of increases in creative innovations at the firm level and the relationship of this to our structural parameters.

4.1 Baseline Results

Our baseline results are presented in Table 2. Our estimating equation is

$$y_f = \alpha m_f + \mathbf{X}'_f \boldsymbol{\beta} + \delta_{i(f)} + \varepsilon_f, \tag{18}$$

where y_f is one of our measures of creative innovations introduced in the previous section (innovation quality, superstar fraction, tail innovation, or generality) for firm f, and m_f is our firm-level measure of openness to disruption, the average age of company CEOs over our sample window. In addition, $\mathbf{X}_f$ is a vector of controls, in this case, firm age, log of employment, log of sales, and log of total number of patents during our time window (we do not have measures of the human capital of the firm's employees).[24] Controlling for firm age is particularly important 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.[25] Finally, ε_f is the error term.

Our baseline sample comprises 279 firms with complete information on CEO and positive patents between 1995 and 2000 (as well as information on firm age, sales, and employment). As noted above, we first exploit only cross-sectional information, so our regressions have one observation per firm,

[23] 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).

[24] Our log employment and log sales variables in the cross-sectional regressions are computed as averages of annual log employment and log sales.

[25] All firms in our baseline sample are in one of 120 four-digit SIC industries.

and are weighted with the total patent count of the firm, so that they put less weight on observations for which our measures of creative innovations are computed from only a few patents. All standard errors are robust against heteroscedasticity. Different columns of Table 2 correspond to our four different measures of creative innovations.

Column 1 shows an economically sizable correlation between CEO age and our measure of innovation quality (average number of citations per patent). The coefficient estimate, -0.278 (standard error = 0.088), is statistically significant at 1% and indicates that companies with a younger CEO have greater innovation quality. We interpret this pattern as evidence that companies that are more open to disruption tend to be the ones producing more creative innovations. The quantitative magnitudes are significant and also plausible. For example, a one-year increase in CEO age is associated with a 0.278 increase in average citations, which is approximately 1.3% of the firm-level weighted mean of our innovation quality variable (20.5).

The pattern of the covariates is 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). Our measures of creative innovations are also uncorrelated with employment and sales, and are largely uncorrelated with the number of patents held by the firm (the exception being a marginally significant relationship for tail innovations). This confirms that our measures of creativity of innovations are quite distinct from the total number of patents.

Column 2 shows a similar relationship with the superstar fraction (-0.300, standard error = 0.141). This also 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 (significant at 1% or less) and economically large relationships with our measures of tail innovations and generality. The implied quantitative magnitudes are a little larger with the superstar fraction and tail innovation measures (a one-year increase in CEO age is associated with, respectively, 2.4% and 5.5% increases relative to weighted sample means in these two measures).

Overall, these results suggest that there is a strong statistical and quantitative relationship between the age of the CEO of a Compustat company and each one of our four measures of creative innovations. We next investigate the robustness of these patterns.

4.2 Robustness

Tables 3 and 4 probe the robustness of our firm-level results in different dimensions. Table 3 looks at the alternative measures of creative innovations (these are a measure of innovation quality using average citations per patent computed using only five years of citations data, a measure of superstar inventors using information on the most highly cited patent of the inventor, the tail innovation index with $p = 0.90$, and the originality index). The results show that the pattern is quite similar to those in Table 2, except that the relationship is no longer statistically significant with the alternative measure of the superstar fraction.

Table 4 looks at several different robustness exercises. Panel A replaces the four-digit SIC dummies with three-digit dummies (a total of 84 in our baseline sample), with effects very similar to our baseline results.

Panel B goes in the opposite direction and enriches the set of controls. In particular, this specification, in addition to the four-digit SIC dummies, 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. For example, CEO age is statistically significant at less than 1% with all of our measures of creative innovations, except for the superstar fraction, for which it is significant at 5%.

Panel C, in addition, adds R&D intensity (R&D to sales ratio) to the previous specification.[26] 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 257 companies). The results are once again very similar to those in our baseline regressions in Table 2.

Panel D uses 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. Nevertheless, the relationship is very similar to this measure as shown in Panel D.

Panels E and F reestimate the specifications in Table 2 for subsamples of high-tech and low-tech firms, where high-tech firms are those in SIC 35 and 36 (industrial and commercial machinery and equipment and computer equipment; and electronic and other electrical equipment and components), and low-tech firms are the rest. This is intended to check whether our results are driven by a subset of firms and whether they are differential between these two subsamples. The results are fairly similar in these two subsamples, except for the superstar fraction variable, which shows a considerably stronger relationship for the low-tech sample.

4.3 Panel Results

We now show that, though naturally much noisier, the correlation between CEO age and creative innovations is present when we exploit within-firm variation in the age of the CEO. We will also document, however, that consistent with our theory, creative innovations start increasing before there is a decline in CEO age.

With this objective in mind, in Table 5 we start with our baseline balanced sample, but now we

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

compute our measures of creativity of innovations at an annual frequency. The covariates we use are also at an annual frequency and include a full set of year dummies. As a first step, in Panel A, we maintain our key right-hand-side variable, average CEO age over the sample period, which is thus held constant across years in this panel. In this table, standard errors are robust for arbitrary heteroscedasticity at the firm level (thus allowing for arbitrary dependence across the observations for the same firm). These specifications are directly comparable to those in Table 2, and indeed, the coefficient estimates and standard errors are very similar (though they are not identical since the covariates are now time-varying).[27]

Panel B extends our sample in two different ways. First, we include firms that were left out of the balanced panel (i.e., firms for which CEO age or patent information is available in some but not all years). Second, with the unbalanced panel, we can now consider a longer sample spanning $1992-2004$ (we cannot go before 1992 because of lack of data on manager age, and we prefer not to go beyond 2004, as this would make the citation window too short and thus our measures much less reliable). The resulting sample has 7111 observations (or 5803 observations with tail innovation, since we lose firm-years when no patent is above the median of the citation distribution). Despite the increase in the number of firms to 1256 (from 279) and the addition of seven more years of data, the results are remarkably similar to those in Panel A and to our baseline estimates.

Panel C allows CEO age to vary across years but also includes firm fixed effects as well as year effects (and, of course, in this case, SIC industry dummies and firm age are dropped). This effectively means that the CEO age variable is being identified from changes in CEOs.[28] Hence, this is a demanding specification investigating whether in years when a firm has a younger CEO, it tends to have more creative innovations, and this motivates our choice of focusing on the $1992-2004$ sample for this exercise. 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 past years.[29] Though these considerations stack the cards against finding a significant relationship between CEO age and creative innovations, the results are generally quite consistent with our cross-sectional estimates from the balanced panel. All of the coefficient estimates in these within-firm regressions, except for generality, have the same sign and are statistically significant as in our baseline results in Table 2. For innovation quality, the magnitude of the estimate is about 12% larger than the specification

[27] The number of observations is now lower in columns 3 and 4 because not all firms have patents with citations above the median (for tail innovations) or with positive citations (for generality) in all years.

[28] 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.

[29] 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 when the CEO is in charge but on patents applied for when the CEO is in charge.

without fixed effects in Panel B (e.g., −0.188 vs. −0.168), whereas for superstar fraction and tail innovations, it is smaller—about 47% to 73% of the magnitude in Panel B.

The current CEO/manager 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 D by including current CEO age and lagged CEO age simultaneously. Our results show that, with all of our measures of creative innovations (except generality), both matter with quantitatively similar magnitudes.

A related question concerns separating the impact of the current CEO from the persistent effects of past innovations—in particular, past creativity may spill over into current creativity in part because patents from the same 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 E show that its coefficient is very far from 1 and the estimates are fairly similar to those in Panel C.[30]

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

Although the results in Panel F suggest that both the sorting and the causal effects of CEO/manager age are important for the creativity of innovations, they do not directly translate into an estimate of the impact of CEO/manager age on creative innovations for a given firm (because changes in CEO age are associated with changes in firm type/corporate culture as well as the firm's prior number of incremental innovations). We next turn to an indirect inference procedure utilizing the structure of our model to obtain an estimate of the size of this causal effect.

[30] 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.

4.4 The Causal Effect of Manager Age on Creative Innovations

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

The (average) impact of a younger manager on the creativity of innovations for a given firm type is $\frac{1}{F(a^*)}\int_0^{a^*}\Lambda\theta_H\bar{q}^a dF(a) = \frac{\Lambda\theta_H\delta}{g+\delta}\frac{\left[1-e^{-(g+\delta)a^*}\right]}{\left[1-e^{-\delta a^*}\right]}$. Because of the sorting of younger managers to firms that are more open to disruption, we cannot read off 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 above.

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 extent of the causal effect of CEO age on creative innovations from a simple indirect inference exercise. 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 $x = 5\%$ which corresponds to the entry rate in our Compustat sample. Finally, we take the parameter α, which determines how rapidly the productivity of incremental innovations declines from Akcigit and Kerr (2015), who estimate a similar parameter from the patent citation distribution.

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

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

In addition to these three regression coefficients, our indirect inference procedure targets three central moments in the data: the average annual growth rate of (real) sales per worker, within-firm coefficient of variation of radical innovations, and the fraction of incremental innovations, measured as fraction of internal patents which mainly build on innovating firm's existing lines (as opposed to innovating on product lines operated by other firms).[31] This implies that we have in total six data moments and six parameters.

Finally, we make two additional assumptions in matching the model to data. First, in the model managers are employed at the product line level, whereas in the data we only observe managers/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/CEO 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 the 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. Table 7 shows the match between the values of these moments in the data and those implied by the model. The model-implied numbers are on the whole very close to the targeted empirical moments. The most important lesson from Table 7 is that the model is quite consistent with reduced-form regression results, including the significant and sizable coefficient on lead CEO age, which is generated by the fact that $\psi > 0$ and is a non-trivial source of creative innovations.

The implied pattern is also visible in Figures 1 and 2, which plot the probability of a creative innovation and the average CEO age as a function of time since switching to high-type. These figures show that firms slowly reduce the average age of their managers after switching to high-type (if at first they are below n^*, they do not need to change their CEO). Correspondingly, they

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

also slowly increase their probability of creative innovations. Because much of this increase in the probability of creative innovations takes place before firms switch to a younger manager, in the reduced-form regressions it will be captured by lead CEO age.

It is also useful to gauge whether, at these estimated parameter values, the model performs well on some other dimensions. One empirical moment we have not used for estimation is the probability of firms switching to younger managers. At the estimated parameter values, 6% of all firms attempt a radical innovation (these are high-type firms with $n \geq n^*$). Consequently, "young" managers (defined as those with $a < a^*$ in Proposition 4) also make up 6% of the population of managers, implying that a^* corresponds to age 43 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 very close to each other, respectively 0.62% and 0.63%.

Using these parameter estimates, we next compute the "causal effect" of manager age on creative innovations. There are several ways in which such a causal effect might be defined. First, we could define the causal effect in a fashion analogous to "treatment effect on the treated," by considering the loss of creative innovation that firms that were previously hiring young managers and pursuing radical innovations would suffer. Second, one could focus on the "average treatment effect," corresponding to the impact of having a younger manager for an average firm in our sample. It is intuitive that these two measures of causal effects will be quite different, since, as just noted, only 6% of firms are attempting radical innovations in our stationary distribution.

For the first, we start with the equilibrium stationary distribution and reshuffle managers only among firms attempting radical innovation (which are high-types with more than n^* incremental innovations and hiring managers younger than a^*), and we repeat this for 13 periods. Because such firms will continue to attempt radical innovation after the reshuffling (since the reshuffling involves only managers younger than a^*), the change in the likelihood of radical innovation of any given firm captures the causal effect of a younger (or older) manager on a firm attempting radical innovation. We quantify this effect by running the same regression of the likelihood of radical innovation on the age of manager for this subsample of firms (corresponding to Panel C) and then comparing it to the reduced-form relationship between these two variables in the model-generated data, -0.211 (as reported in Table 7).[32] The resulting causal effect is estimated as -0.040. This effect is thus considerably smaller than the reduced-form regression coefficient of -0.211, though it should not be directly compared to this number, which applies to the entire sample, while the causal effect of -0.040 is only for 6% of the entire sample. To obtain a causal effect estimate more comparable to

[32] We use the regression coefficient obtained from model-generated data rather than the regression coefficient from Table 5, -0.188, since this will be compared to numbers also obtained from model-generated data.

the reduced-form relationship, we next turn to the average treatment effect.

For this second exercise, we again start with the stationary distribution and reshuffle managers randomly for 13 periods. We assume that after the reshuffling, each firm will pursue the same innovation strategy.[33] We then use the data generated by this thought experiment to run a regression of the likelihood of a radical innovation on manager age for the entire sample of firms (including both low-type firms and high-type firms not attempting a radical innovation). This exercise yields an average causal effect of -0.003, and thus accounts approximately for 1.5% ($\simeq 0.003/0.211$) of the relationship between CEO age and creative innovations. The rest of this relationship is explained by sorting effects—because it is high-type firms that are hiring younger managers.

It is intuitive that the first estimate of the causal effect is much larger than the second, because it explicitly focuses on the small subsample of firms attempting a radical innovation. But even in this case, especially once we take into account that this causal effect applies only to 6% of the sample of firms that are attempting a radical innovation, much of the association between manager age and creative innovations is accounted for by the sorting of younger managers to firms that are more open to disruption.

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

4.5 Inventor Age and Creativity of Innovations

We next turn to patent-level regressions to investigate the relationship between the age of inventor—defined as any inventor listed in our patent data—and our various measures of creativity of innovations. Though in our theoretical model there is no distinction between managers and inventors, this distinction is of course 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 effect of manager/inventor age on the creativity of innovations once we control for the type of characteristics of the firm.

[33] 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, but estimates 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 unless there is a very large change in the age of the manager assigned to them.

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

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

$$y_{ift} = \phi I_{ift} + \alpha m_{ft} + \mathbf{X}'_{ift}\boldsymbol{\beta} + \delta_f + \gamma_i + d_t + \varepsilon_{ift}. \tag{19}$$

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 also control for a full set of year effects, denoted by d_t, and ε_{ift} is the error term.[34]

The results from the estimation of (19) are reported in Table 8. In Panel A we focus on a specification similar to the regressions with firm fixed effects reported in Table 5. This is useful for showing that this different frame still replicates the results showing the impact of CEO age on creativity of innovations. In particular, Panel A focuses on Compustat firms for the period $1992-2004$ and includes the same set of controls as in Table 5 Panel C (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 C of Table 5, though a little smaller. In column 1, for instance, we see an estimate of -0.119 (standard error $= 0.038$) compared to -0.188 in Table 5. We cannot define our measure of the superstar fraction and tail innovations in these patent-level regressions. We can, however, look at a patent-level measure related to tail innovations, a dummy for the patent in question being above the pth percentile of the citation distribution. We 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 negatively correlated with CEO age, though only marginally significantly in these specifications.[35]

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

[35] For completeness, we also show results with the generality index, even though the results in Table 5 already indicated that, with firm fixed effects included, there is no longer a significant relationship between CEO age and

Panel B goes in the other direction and reports the estimates of a model that controls for inventor characteristics and looks at the impact of inventor age, without controlling for CEO age, for the same sample as in Panel A (thus restricting it to firms with information on CEO age). As with all of the other models reported in this table, in Panel B we control for a full set of dummies for the maximum number of patents of any inventor associated with the patent in question has over our sample period;[36] 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.[37] The inclusion of this rich set of dummy variables enables us to compare inventors of similar productivity. It thus approximates a model that includes a full set of inventor dummies.[38] The results show that there is a strong relationship between inventor age and the creativity of innovations. For example, in column 1, the coefficient estimate on inventor age is -0.234 (standard error = 0.026), about twice as large as the CEO age estimate in Panel A.

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

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

Panel E provides our main results in this subsection. It returns to the Compustat sample over the period 1992-2002 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 -0.111 (standard error = 0.038) and that on inventor age is -0.235 (standard error = 0.027);

the generality index, and this lack of relationship persists for all of the estimates we report in Table 8 (and for this reason, though we do show them for completeness, we will not discuss them in detail).

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

[37] This corresponds to 374 separate technology classes and is roughly at the same level of disaggregation as the SIC dummies we used in the firm-level analysis in Tables 2-4.

[38] We cannot include a full set of inventor fixed effects directly because inventor age would not be identified in this case since we also have a full set of year dummies.

these are very close to the estimates in Panels A and B, respectively. The pattern is similar in the other columns (except again for generality).

These results provide further evidence that the relationship between manager/CEO age and the creativity of innovations in the data reflects an important dimension of sorting. In particular, firms appear to make several associated changes—in top management and innovation teams—around the same time they change their portfolio of innovation and their innovation strategy (and perhaps their "corporate culture"). Reflecting this sorting, the estimated magnitudes linking CEO age to our indices of creative innovations are smaller in Table 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. In particular, we estimate a regression similar to equation (19) 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 firm fixed effects are again controlled for. 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 (though only marginally significant) relationship, suggest that younger CEOs tend to hire younger inventors, indirectly corroborating the sorting effect emphasized in our theoretical model.[39]

4.6 Stock of Knowledge, Opportunity Cost and Creativity of Innovations

Finally, Table 10 turns to an investigation of some additional implications of our approach already highlighted in our theoretical model (in particular, Proposition 5). We noted there that we may expect openness to disruption to be more important for companies that are technologically more advanced (as measured by the number of patents), but also that companies that have more to lose (because of the greater opportunity cost of disruption in terms of other profitable activities) may shy away from disruptive creative innovations. The firm-level data enable us to look at 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). 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

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

sales and log patent counts are positively correlated (the weighted correlation between the two variables is 0.7 in our sample), thus stacking the cards against finding an informative set of results.

Nevertheless, Table 10, which uses the same unbalanced sample with annual observations as in Table 5 Panel C, provides some evidence that our theoretical expectations are partially 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.[40] These results thus provide some support for the hypothesis that the stock of knowledge of the company and opportunity cost effects might be present and might in fact be quite important (at least quantitatively at this correlational level).

5 Cross-Country Correlations

In this section, we provide evidence that the firm-level relationship between manager age and creativity of innovations appears to aggregate up to the national level. In particular, we document that there is a cross-country relationship between manager age and creativity of innovations. Moreover, at the cross-country level, we can also use other indices potentially proxying for openness to disruption, which also show similar results, thus partially corroborating our interpretation of the manager age variables in our firm-level and cross-country empirical work.

The interpretation of the cross-country relationships should be somewhat different than the firm-level ones. At the country level, manager age, like our other measures of openness to disruption presented below, is likely to have its impact on the creativity of innovations not just because of its association with— and because of its impact on—firm-level innovation strategies, but also through economy-wide institutions, attitudes and values of the society. This suggests that the quantitative magnitudes of the relationships might be somewhat larger at the country level than at the firm level.

Our main cross-country results are presented in Table 11, which reports OLS regressions of the following form:

$$y_c = \alpha I_c + \mathbf{X}_c' \boldsymbol{\beta} + \varepsilon_c, \tag{20}$$

where y_c is one of our measures of creative innovations (innovation quality, superstar fraction, tail innovation, or generality) for country c, I_c denotes one of our measures of openness to disruption (the individualism index, the uncertainty avoidance index, or average manager age), $\mathbf{X}_c$ is a vector of controls (including average log real GDP per capita of the country, average years of secondary schooling and log of total patents of the country during this time period), and finally, ε_c is an

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

error term.[41] The coefficient of interest is α, which will reveal whether there is a cross-country correlation between our measures of openness to disruption and the creativity of innovations.

All regressions in Table 11 include one observation per country. As with our firm-level specifications, these regressions are weighted using the total number of patents as weights, which is again motivated by the fact that countries with more patents are both more important for their contribution to creative innovations and have much more precisely estimated measures for our key variables (see Appendix Table A1 for the distribution of total number of patents across countries). All standard errors continue to be robust against heteroscedasticity.

Panel A of Table 11 focuses on our measure of manager age (which is available for 37 countries). The patterns are very similar to those we obtained in the firm-level analysis, and show a strong correlation between average manager age and all four of our measures of creative innovations. For example, in column 1, the estimate of α is -0.484 (standard error = 0.225). We also see that log GDP per capita and average years of secondary schooling are not significant correlates of the creativity of innovations, while log patent count is significant and indicates that countries that have more patents also tend to have more creative innovations. Consistent with the caveat about the interpretation of the cross-country results, the quantitative magnitudes are somewhat larger than the firm-level ones: a one-year change in manager age increases average citations by 0.48 (3.3% compared to its mean of 14.5), the superstar fraction by 0.96 (14.4% relative to its mean), tail innovations by 0.23 (11.7% relative to its mean) and generality by 0.28 (1.3% relative to its mean). These effects are about 2 to 5 times larger than the firm-level estimates presented above.[42]

Panel B is similar to Panel A, except that it uses Hofstede's individualism index (this increases the sample from 37 to 50). The results are very similar to those using average manager age, and the quantitative magnitudes of the correlation between individualism and innovation quality are again sizable and somewhat larger than those in Panel A.[43]

Panel C has exactly the same structure, except that the right-hand-side variable is Hofstede's uncertainty avoidance index. The patterns are very similar and generally even more precisely esti-

[41] An additional covariate that might be useful to control for would be the average educational attainment of managers in a country. Though this number is available in the World Bank dataset that Gennaioli et al. (2013) use, there is very little overlap between this developing country sample and ours. We have instead experimented with controlling for the average education of the managers of the companies we have used for compiling our average manager age variable. This has no effect on the results reported here and is omitted to save space. The details are available upon request from the authors.

[42] If, instead, we look at the quantitative implications of moving from the 75th percentile of the manager age distribution to the 25th percentile, the magnitudes are more similar to the firm-level estimates. For example, moving from the country at the 25th percentile of average manager age in our sample to the 75th percentile (from 51.5 to 54.3) reduces our measure of innovation quality by 9.4% relative to the sample mean (14.5).

[43] For example, moving from the country at the 25th percentile of individualism in our sample to those at the 75th percentile (from 0.19 to 0.73) increases our measure of innovation quality by 19% relative to its weighted sample mean (14.5). Using the same metric for quantitative magnitude for the average manager age gives an increase in innovation quality by 9.4% relative to the sample mean (14.5).

mated (though, of course, they are now negative, since greater uncertainty avoidance corresponds to less openness to disruption). The quantitative magnitudes are similar to those in Panel B.[44]

Tables 12 and 13 probe the robustness of these cross-country relationships. Table 12 looks at various alternative measures of creative innovations (which we also investigate at the firm level). These are average citations per patent but now constructed using only a five-year window (so that we do not have to rely on the correction factors); an alternative measure of the superstar fraction of patents but now computed using information on the most highly cited patent to the inventor (rather than lifetime average citations); the tail innovation index computed with $p = 0.90$ (instead of $p = 0.99$); and the originality index mentioned above. The results in all cases are similar to the baseline (though weaker and not statistically significant with the alternative measure of superstar fraction).

Table 13 , on the other hand, investigates whether these results can be explained by the fact that R&D intensity (defined as total R&D spending divided by GDP at the country level) differs across countries. Our results largely might be reflecting the fact that some countries invest more in R&D and as a result generate more creative innovations. However, in our sample R&D intensity is not systematically related to individualism, uncertainty avoidance, or average manager age. Moreover, Table 13 shows that controlling for variation in R&D intensity does not change the basic correlations in our sample. The parameter estimates do change in some cases, particularly with the individualism variable, but the association between our measures of openness to disruption and creativity of innovations always remains highly significant.[45]

6 Conclusion

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

[44] We do not run regressions including multiple indices at the same time, since we believe this type of horse race would not be particularly informative. Instead, we interpret each of these variables as a proxy for the same underlying tendency for openness to innovation, new practices and ideas.

[45] We also experimented with using cross-country differences in demographics to instrument for average manager age differences. Though these results corroborate the patterns shown here, we do not report them both because demographics could have a direct effect on the creativity of innovations, invalidating the exclusion restriction of such a strategy, and because we view the cross-country results as additional evidence rather than as our main empirical focus.

We provided a simple model drawing a clear distinction between radical (more creative) innovations and incremental innovations, whereby the former combines ideas from several different lines of research and creates more significant advances (and contributions to knowledge). These advances can be discouraged or even stopped, either through pecuniary or non-pecuniary means, preventing radical innovations directly or discouraging cross-fertilization of ideas from different fields.

The bulk of our paper provides illustrative cross-country and firm-level correlations consistent with the role of openness to disruption. We use several measures to proxy for creative innovations. These include our proxy for innovation quality, which is the average number of citations per patent; two indices for creativity of innovations, which are the fraction of superstar innovators and the likelihood of a very high number of citations (in particular, tail citations relative to median citations); and the generality index.

Our main proxy for openness to disruption is the age of the CEO or top management of the company (or the average age of the CEO and CFO of the top 25 publicly listed companies in a country). This variable is motivated based on the idea that only companies or societies open to such disruption will allow the young to rise up within the hierarchy. This is the only variable we have available as a proxy for openness to disruption at the firm level. At the country level, we augment this variable with the popular indices for individualism and uncertainty avoidance based on the work by the Dutch social scientist Geert Hofstede.

Using these proxies, at the firm, patent and country level, we find fairly consistent and robust correlations between openness to disruption and creative innovations. We also show that these relationships are generally robust. They do not, however, correspond to the causal effect of CEO (or manager) age on creative innovations because, as highlighted by our theoretical model, younger managers tend to be employed by firms that are more open to disruption and more creative. A simple indirect inference exercise using the structure of our model suggests that most of the empirical relationship between CEO age and creative innovations is due to these sorting effects, and the causal impact of CEO age is quite small.

Finally, our theoretical model further suggests that the impact of openness to disruption should be larger for companies that are technologically more advanced (closer to the technology frontier) and smaller for companies that have a greater opportunity cost of disruptive innovation. The empirical patterns in our firm-level data support this prediction.

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

Appendix: Omitted Proofs

Proof of Proposition 2. We conjecture that the value function for low-type firms takes the form in (10). Substituting this conjecture into (7), we get

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

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

$$rAq_j = \pi q_j - \tau A q_j,$$

and

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

Solving these equations for A and $B(n)$, while taking $\tilde{B}(n)$ as given and to be determined in Proposition 3, completes the proof. ■

Proof of Proposition 3. Following the same steps, we conjecture that the value function for high-type firms takes the form in (12), and substitute this into (8) to get

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

which implies

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

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

$$\begin{aligned} (r-g+\tau)\tilde{B}(n) &= \psi\Lambda\theta_H\left[A(1+\eta) + \tilde{B}(0)\right] \\ &\quad + \max\left\{\begin{array}{c} \xi\left[A\eta\alpha^{n+1} + \tilde{B}(n+1) - \tilde{B}(n)\right]; \\ \Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A + \tilde{B}(0)\right] \end{array}\right\}. \end{aligned} \tag{21}$$

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

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

Under the hypothetical scenario where the max operator in (21) always picks the first term, we have $\tilde{B}(n) = \hat{B}(n)$. Collecting terms, we can write

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

where $\beta = \frac{\xi}{(r-g+\tau+\xi)}$ and $\hat{\psi} = \psi\Lambda\theta_H\left[A(1+\eta) + \tilde{B}(0)\right]$. From standard dynamic programming arguments (e.g., Theorem 4.7 in Stokey and Lucas, 1989), $\hat{B}(n)$ is strictly decreasing and limits to $\frac{\hat{\psi}}{r-g+\tau}$. Now note that if $n^* = \infty$ (meaning that incremental innovations were always optimal), then we would have $\tilde{B}(n) = \hat{B}(n)$.

The other option in the max operator, $\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A + \tilde{B}(0)\right]$, does not depend on n and is strictly positive, which implies that switching to radical innovation for n sufficiently high would yield $\tilde{B}(n) = \frac{\hat{\psi}+\Lambda\theta_H\bar{q}^{a^*}\left[(1+\eta)A+\tilde{B}(0)\right]}{r-g+\tau} > \frac{\hat{\psi}}{r-g+\tau}$. Hence, there must exist n^* such that firms with $n < n^*$ undertake incremental innovation and switch to radical innovation at n^*. The expression for n^* follows by equating the value of pursuing radical and incremental innovations at n' and setting n^* as the smallest integer greater than n'. ■

Proof of Proposition 5. Recall that

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

where we have written 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).

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

$$V_H\left(q_{n,t}^*+\eta_{n+1,t}(q_{n,t}^*),n+1\right) - V_H\left(q_{n,t}^*,n\right) = \frac{\Lambda\theta_H\bar{q}^a}{\xi}\mathbb{E}V_H(t), \tag{22}$$

and at this value, we also have

$$(r+\tau)V_H\left(q_{n,t}^*,n\right) - \dot{V}_H\left(q_{n,t}^*,n\right) = \pi q_{n,t}^* + \Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t) + \psi\Lambda\theta_H\mathbb{E}V_H(t). \tag{23}$$

Now we will consider two alternative cases:

Case 1:

$$q_{n+1,t}^* \geq q_{n,t}^* + \eta_{n+1,t}(q_{n,t}^*). \tag{24}$$

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

Under this case, we have

$$
\begin{aligned}
& (r+\tau)V_H\left(q^*_{n,t}+\eta_{n+1,t}(q^*_{n,t}),n+1\right)-\dot{V}_H\left(q^*_{n,t}+\eta_{n+1,t}(q^*_{n,t}),n+1\right) \\
= \quad & \pi q^*_{n,t}+\pi\eta_{n+1,t}(q^*_{n,t})+\Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t)+\psi\Lambda\theta_H\mathbb{E}V_H(t).
\end{aligned} \tag{25}
$$

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 (22) with respect to time, we have

$$
\begin{aligned}
\dot{V}_H\left(q^*_{n,t}+\eta_{n+1,t}(q^*_{n,t}),n+1\right)-\dot{V}_H\left(q^*_{n,t},n\right) &= \frac{\Lambda\theta_H\bar{q}^a}{\xi}\partial\mathbb{E}V_H(t)/\partial t \\
&= \frac{\Lambda\theta_H\bar{q}^a}{\xi}g\mathbb{E}V_H(t),
\end{aligned} \tag{26}
$$

where, in the second line, we have used the fact that in a stationary equilibrium $\mathbb{E}V_H(t)$ grows at the rate g. Subtracting (23) from (25) and using (26), we obtain:

$$
(r+\tau)[V_H\left(q^*_{n,t}+\eta_{n+1,t}(q^*_{n,t}),n+1\right)-V_H\left(q^*_{n,t},n\right)]=\pi\eta_{n+1,t}(q^*_{n,t})+\frac{\Lambda\theta_H\bar{q}^a}{\xi}g\mathbb{E}V_H(t). \tag{27}
$$

Then, combining (22) and (27) we can derive

$$
\pi\eta_{n+1,t}(q^*_{n,t})=\frac{r-g+\tau}{\xi}\Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t). \tag{28}
$$

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

Now let us define

$$
v_t\equiv\frac{r-g+\tau}{\pi\xi}\Lambda\theta_H\bar{q}^a\mathbb{E}V_H(t), \tag{29}
$$

which is independent of both q and n. Using (29) equation (28) can be written as

$$
[\kappa\bar{q}_t+(1-\kappa)q^*_{n,t}]\eta\alpha^{n+1}=v_t, \tag{30}
$$

or

$$
q^*_{n,t}=\frac{v_t/\eta\alpha^{n+1}-\kappa\bar{q}_t}{1-\kappa}. \tag{31}
$$

This equation immediately 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 (24) indeed applies. For this reason, note that from (30) written for $n+2$ incremental innovations, we have

$$
q^*_{n+1,t}=\frac{v_t/\eta\alpha^{n+2}-\kappa\bar{q}_t}{1-\kappa}. \tag{32}
$$

Combining equations (31) and (32), we obtain that (24) is satisfied if

$$
(1-\kappa)\eta\alpha^{n+2}+\alpha\leq 1. \tag{33}
$$

Thus whenever (33) 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 (33) holds, the same result applies.

Case 2:

$$q_{n+1,t}^* - \eta_{n+1,t}(q_{n,t}^*) < q_{n,t}^*. \tag{34}$$

This implies that if a high-type firm is indifferent between radical and incremental innovation at $n+1^{st}$ prior incremental innovations at time t, then it would have preferred to switch to radical innovation at n^{th} prior incremental innovations. This condition is clearly the complement of (24).

In this case, start with $q_{n+1,t}^*$, which satisfies (25). Under condition (34), $q_{n,t}^*$ satisfies (23), so we again arrive at (22), (28) and (31). But then from (31) $q_{n,t}^*$ 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 (33) holds. Note that the same expressions for $q_{n+1,t}^*$ as in (32) again applies under Case 2. Thus the condition for (34) to be satisfied, with an identical argument, is

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

which is indeed the complement of (33).

Consequently, regardless of whether (33) or its converse holds, equation (31) applies, and $q_{n,t}^*$ is increasing in n (or equivalently, $n_t^*(q)$ is increasing in q). This completes the proof of the proposition.

■

References

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

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

Adorno, T. W., E. Frenkel-Brunswik, D. J. Levinson, and R. N. Sanford (1950). *The Authoritarian Personality.* Harpers.

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

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

Alesina, A., P. Giuliano, and N. Nunn (2011). Fertility and the Plough. *American Economic Review,* 101 (3), 499-503.

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

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

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

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

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

Barro, R. and J. W. Lee (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104, 184–198.

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

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

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

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

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

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

Cyert, R. M. and G. J. March (1963). *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

Durkheim, E. (1933). *The Division of Labor*. New York: Macmillan.

Fernandez, R. and A. Fogli (2009). Culture: An Empirical Investigation of Beliefs, Work and Fertility. *American Economic Journal: Macroeconomics*, 1 (1), 146-177.

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

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

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

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

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

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

Gennaioli, N. and R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2013). Human Capital and Regional Development, *Quarterly Journal of Economics*, 128(1), 105–164.

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

Greif, A. (1994). Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. *Journal of Political Economy*, 102(5), 912-950.

Guiso, L., P. Sapienza, and L. Zingales (2010). Civic Capital as the Missing Link. National Bureau of Economic Research Working Paper 15845.

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

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

Hofstede, G., G. J. Hofstede, and M. Minkov (2010). *Cultures and Organizations: Software for the Mind*. McGraw-Hill.

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

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

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

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

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

Kaplan, S. N., M. M. Klebanov, and M. Sorensen (2012). Which CEO Characteristics and Abilities Matter? *Journal of Finance,* 67(3), 973-1007.

Kirzner, I. M. (1997). Entrepreneurial Discovery and the Competitive Market Process: An Austrian Approach. *Journal of Economic Literature,* 35 (1), 60-85.

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

Kluckhohn, F. R. and F. L. Strodtbeck (1961). *Variations in Value Orientations.* Evanston, IL: Row Peterson.

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

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

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

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

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

McClelland, D. (1961). *The Achieving Society.* The Free Press, New York.

Parsons, T. (1949). *Essays in Sociological Theory; Pure and Applied.* Free Press.

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

Sarada, and Tocoian , O. (2013). The Success of Entrepreneurial Networks: Evidence from Brazil. Duke University Working Paper.

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

Schwartz, S. H. (1994). *Beyond Individualism/Collectivism: New Cultural Dimensions of Values.* Sage Publications, Inc.

Schwartz, S. H. and W. Bilsky (1990). Toward a Theory of the Universal Content and Structure of Values: Extensions and Cross-cultural Replications. *Journal of Personality and Social Psychology,* 58(5), 878.

Shane, S. (1993). Cultural Influences on National Rates of Innovation. *Journal of Business Venturing,* 8(1), 59-73.

Shane, S. (1995). Is the Independent Entrepreneurial Firm a Valuable Organizational Form? In *Academy of Management Proceedings*, 110-113. Academy of Management.

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

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

Tabellini, G. (2008a). Institutions and Culture. *Journal of European Economic Association*, 6(2), 255-294.

Tabellini, G. (2008b). The Scope of Cooperation: Values and Incentives. *Quarterly Journal of Economics*, 123(3), 905-950.

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

Triandis, H. C. (1995). *Individualism & Collectivism.* Westview Press.

Walter, F., and S. Scheibe (2013). A Literature Review and Emotion-based Model of Age and Leadership: New Directions for the Trait Approach. *Leadership Quarterly,* 24(6), 882-901.

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

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

Winslow, E. K. and G. T. Solomon (1987). Entrepreneurs Are More Than Non-Conformists: They Are Mildly Sociopathic. *Journal of Creative Behavior,* 21(3), 202-213.

Table 1: Summary Statistics

Panel A: Descriptive Statistics

Variable	Observations	Mean	Standard Deviation
Balanced Firm Sample (Firm Averages, 1995-2000)			
CEO age	279	55.3	6.47
average manager age	279	52.3	4.32
innovation quality	279	20.5	8.76
superstar fraction	279	12.3	10.1
tail innovation	279	2.72	2.56
generality	279	21.5	5.53
log patents	279	5.86	1.51
log employment	279	3.84	1.38
log sale	279	4.34	1.47
firm age	279	37.3	14.4
R&D intensity	257	8.52	17.0
Unbalanced Firm Sample (Annual Firm Observations, 1992-2004)			
CEO age	7111	55.3	6.84
average manager age	7111	52.3	4.38
innovation quality	7111	15.9	10.9
superstar fraction	7111	9.91	10.7
tail innovation	5803	3.41	5.42
generality	6232	18.5	9.96
log patents	7111	5.61	1.60
log employment	7111	3.71	1.51
log sale	7111	4.12	1.61
firm age	7111	35.1	16.3
Cross-Country Sample (Country Averages, 1995-2000)			
individualism	50	.813	.263
uncertainty aversion	50	.492	.195
average manager age	37	56.1	2.98
innovation quality	50	14.5	3.26
superstar fraction	50	6.68	3.65
tail innovation	50	1.92	.945
generality	50	21.0	1.81
log patents	50	10.5	1.52
log income per capita	50	10.3	.305
secondary years of schooling	50	4.84	.827
R&D intensity	44	2.59	.363

- Table 1 continued on next page -

Panel B: Correlation Matrix of Firm-Level Innovation Variables

	innovation quality	superstar fraction	tail innovation	generality
innovation quality	1.000			
superstar fraction	0.925	1.000		
tail innovation	0.893	0.829	1.000	
generality	-0.177	-0.204	-0.145	1.000

Panel C: Correlation Matrix of Cross-Country Innovation Variables

	innovation quality	superstar fraction	tail innovation	generality
innovation quality	1.000			
superstar fraction	0.932	1.000		
tail innovation	0.945	0.990	1.000	
generality	0.902	0.880	0.906	1.000

Panel D: Correlation Matrix of Openness to Disruption Variables

	individualism	uncertainty aversion	average manager age
individualism	1.000		
uncertainty aversion	-0.884	1.000	
average manager age	-0.770	0.844	1.000

Notes: All statistics in this table are weighted by the number of patents (of the country or the firm). Individualism and uncertainty aversion are Hofstede's indices of national cultures (and are normalized to lie between 0 and 1), and country average manager age is the average manager of CEOs and CFOs of up to the 25 largest firms in the country. Innovation quality is the average number of citations per patent (using the truncation correction weights devised by Hall, Jaffe, and Trajtenberg, 2001); superstar fraction is the fraction of patents accounted for by superstar researchers (those above the 95th percentile of the citation distribution); tail innovation is the fraction of patents of a country or firm above the 99th percentile of the citation distribution divided by the fraction of patents above the median of the distribution; and generality index measures the dispersion of citations received across two-digit IPC technology classes. Log income per capita at the country level, and log employment, log sales at the firm level are computed as the average of, respectively, annual log income per capita, log employment and log sale between 1995 and 2000. 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 balanced firm panel is the sample of firms from Compustat with complete data on CEO age, employment, sales, and firm age and positive patents in each year between 1995 and 2000. The unbalanced firm panel is a sample of firms from Compustat with at least one year of complete data between 1992 and 2002. See text for the definition of other variables and further details.

Table 2: Baseline Firm-Level Regressions

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
CEO age	-0.278	-0.300	-0.151	-0.183
	(0.088)	(0.141)	(0.054)	(0.055)
firm age	-0.219	-0.238	-0.063	0.029
	(0.078)	(0.106)	(0.029)	(0.046)
log employment	-1.599	-4.813	-0.908	-4.574
	(1.937)	(3.376)	(0.793)	(1.500)
log sales	1.833	5.215	0.743	4.421
	(1.425)	(2.645)	(0.650)	(1.331)
log patent	1.073	0.093	0.662	-0.696
	(0.769)	(1.336)	(0.356)	(0.633)
R^2	0.88	0.81	0.79	0.83
N	279	279	279	279

Notes: Weighted cross-sectional regressions with total number of patents as weights. The sample is the balanced firm panel and each observation is the sample average between 1995-2000 as described in the notes to Table 1. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality (the last three are multiplied by 100 to ease legibility). In addition, all regressions control for a full set of dummies for four-digit SIC industries. See text and notes to Table 1 for variable definitions. Robust standard errors are in parentheses.

Table 3: Firm-Level Regressions (Alternative Measures)

	Innovation Quality (5 years)	Superstar Fraction (Best Patent)	Tail Innovation (90/50)	Originality
CEO age	-0.129	-0.497	-0.299	-0.285
	(0.041)	(0.332)	(0.094)	(0.075)
R^2	0.87	0.87	0.83	0.87
N	279	279	279	279

Notes: Weighted cross-sectional regressions with total number of patents as weights. The sample is the balanced firm panel and each observation is the sample average between 1995-2000 as described in the notes to Table 1. The dependent variables are alternative measures of innovation quality (computed over the next five years), superstar fraction (with superstars defined according to the best patent), tail innovation (with share of the patents of the firm among all the patents above the 90th percentile of the citation distribution in the numerator), and the originality index (the last three are multiplied by 100 to ease legibility). All regressions control for firm age, log employment, log sales, log total patents, and a full set of dummies for four-digit SIC industries. See text and notes to Table 1 for variable definitions. Robust standard errors are in parentheses.

Table 4: Firm-Level Regressions (Robustness)

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
Panel A: With SIC3 Dummies				
CEO age	-0.257	-0.284	-0.126	-0.086
	(0.070)	(0.123)	(0.050)	(0.091)
R^2	0.77	0.72	0.64	0.70
N	279	279	279	279
Panel B: With Additional Controls				
CEO age	-0.270	-0.282	-0.150	-0.194
	(0.090)	(0.140)	(0.052)	(0.054)
R^2	0.88	0.82	0.79	0.83
N	279	279	279	279
Panel C: With Additional Controls Plus R&D Intensity				
CEO age	-0.258	-0.295	-0.142	-0.184
	(0.088)	(0.149)	(0.048)	(0.053)
R^2	0.89	0.82	0.81	0.84
N	257	257	257	257
Panel D: With Average Manager Age				
average manager age	-0.418	-0.467	-0.224	-0.339
	(0.163)	(0.206)	(0.094)	(0.084)
R^2	0.87	0.81	0.77	0.83
N	279	279	279	279
Panel E: High-Tech Subsample				
CEO age	-0.227	-0.191	-0.145	-0.189
	(0.068)	(0.157)	(0.045)	(0.043)
R^2	0.92	0.84	0.86	0.81
N	87	87	87	87
Panel F: Low-Tech Subsample				
CEO age	-0.439	-0.704	-0.143	-0.153
	(0.200)	(0.252)	(0.085)	(0.146)
R^2	0.85	0.82	0.72	0.86
N	192	192	192	192

Notes: Weighted cross-sectional regressions with total number of patents as weights. The sample is the balanced firm panel and each of the ratios is the sample average 1995-2000 as described in the notes to Table 1.The dependent variables are innovation quality, superstar fraction, tail innovation, and generality (the last three are multiplied by 100 to ease legibility). Each panel is for a different specification. Unless otherwise stated, all regressions control for firm age, log employment, log sales, log total patents, and four-digit SIC dummies (see text and notes to Table 1 for variable definitions). Robust standard errors are in parentheses. Panel A controls for three-digit SIC dummies instead of the four-digit dummies. Panel B adds to the specification of Table 2 profitability (profit over sales), indebtedness (debt over sales) and log physical capital. Panel C adds to the specification of Panel B R&D intensity (R&D expenditure over sales). Panel D uses average manager age instead of CEO age. Panels E and F are for the high-tech and low-tech subsamples. High-tech sample 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 sample includes the rest.

Table 5: Firm-Level Panel Regressions

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
Panel A: Average CEO Age (No Fixed Effects), Balanced Firm Sample, 1995-2000				
average CEO age	-0.227	-0.336	-0.132	-0.183
	(0.068)	(0.103)	(0.041)	(0.044)
R^2	0.70	0.69	0.47	0.75
N	1,674	1,674	1,594	1,655
Panel B: Average CEO Age (No Fixed Effects), Unbalanced Firm Sample, 1992-2004				
average CEO age	-0.168	-0.319	-0.104	-0.171
	(0.075)	(0.133)	(0.045)	(0.044)
R^2	0.66	0.54	0.31	0.77
N	7,111	7,111	5,803	6,232
Panel C: CEO Age (Fixed Effects), Unbalanced Firm Sample, 1992-2004				
CEO age	-0.188	-0.149	-0.076	0.036
	(0.044)	(0.051)	(0.023)	(0.029)
R^2	0.78	0.80	0.44	0.85
N	7,111	7,111	5,803	6,232
Panel D: CEO Age and Lagged CEO Age (Fixed Effects), Unbalanced Firm Sample, 1993-2004				
CEO age	-0.131	-0.098	-0.052	0.031
	(0.041)	(0.039)	(0.023)	(0.026)
lagged CEO age	-0.123	-0.100	-0.055	0.020
	(0.051)	(0.049)	(0.029)	(0.035)
R^2	0.80	0.81	0.46	0.85
N	5,407	5,407	4,562	4,780
Panel E: CEO Age and Lagged Dependent Var (Fixed Effects), Unbalanced Firm Sample, 1993-2004				
CEO age	-0.096	-0.075	-0.065	0.037
	(0.026)	(0.030)	(0.019)	(0.024)
lagged dependent variable	0.472	0.452	0.194	0.200
	(0.034)	(0.046)	(0.051)	(0.042)
R^2	0.86	0.86	0.46	0.86
N	5,985	5,985	4,772	5,207
Panel F: CEO Age and Lead CEO Age (Fixed Effects), Unbalanced Firm Sample, 1992-2003				
CEO age	-0.113	-0.084	-0.042	0.042
	(0.042)	(0.048)	(0.019)	(0.029)
lead CEO age	-0.125	-0.109	-0.043	-0.007
	(0.049)	(0.044)	(0.022)	(0.028)
R^2	0.78	0.81	0.48	0.85
N	5,409	5,409	4,849	5,097

Notes: Weighted firm-level panel regressions with annual observations with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality (the last three are multiplied by 100 to ease legibility). Robust standard errors clustered at the firm level are in parentheses. Panel A is for our balanced firm sample 1995-2000, and controls for firm age, log employment, log sales, log patents, a full set of four-digit SIC dummies, and year dummies (and thus no firm dummies), and the key right-hand side variable is average CEO age (constant over time). Panel B is identical to Panel A except that the sample is extended to the unbalanced firm panel 1992-2002. In Panel C, the key right-hand side variable is CEO age (in that year), and the regression also includes a full set of firm fixed effects (and thus firm age and the four-digit SIC dummies are no longer included). Panel D is identical to Panel C except that it also includes a one year lag of CEO age as well as current CEO age, and Panel E is identical to Panel C except that it also includes a one year lag of the dependent variable on the right-hand side. See text and notes to Table 1 for variable definitions.

Table 6: Structural Parameters

Parameter	*Description*	*Identification*
	External Calibration	
$x = 0.05$	Entry rate	Compustat sample
$\rho = 0.02$	Discount rate	Standard value
$\delta = 0.04$	Manager death rate	Compustat sample
$\alpha = 0.93$	Reduction rate of innovation size	Akcigit and Kerr (2015)
	Indirect Inference	
$\psi = 10.2$	Baseline radical innovation rate for high type	Estimate
$\Lambda\theta_H = 0.005$	High-type innovation parameter	Estimate
$\varphi = 0.149$	Transition rate from low type to high type	Estimate
$\xi = 0.031$	Incremental innovation rate	Estimate
$\eta = 0.449$	Initial innovation size	Estimate
$\zeta = 0.254$	Probability of high-type entrant	Estimate

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

Table 7: Empirical and Model-Generated Moments

Target	*U.S. Data*	*Model*
Current manager age coefficient of Table 5 Panel C	−0.188	−0.211
Lead manager age coefficient of Table 5 Panel F	−0.125	−0.129
Current manager age coefficient of Table 5 Panel F	−0.113	−0.111
Annual growth rate	5.75%	5.39%
Within-firm coefficient of variation of radical innovations	1.99	2.17
Fraction of internal patents	21.5%	23.8%

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

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	-0.119	-0.314	-1.239	0.028
	(0.038)	(0.132)	(0.413)	(0.025)
R^2	0.11	0.03	0.07	0.11
N	316,516	316,516	316,516	263,641
Panel B: Inventor Age, Unbalanced Firm Sample, 1992-2004				
inventor age	-0.234	-0.440	-2.883	-0.019
	(0.026)	(0.121)	(0.321)	(0.022)
R^2	0.14	0.03	0.09	0.15
N	316,516	316,516	316,516	263,641
Panel C: Inventor Age, Extended Sample, 1985-2004				
inventor age	-0.226	-0.377	-2.842	-0.017
	(0.022)	(0.075)	(0.293)	(0.017)
R^2	0.16	0.05	0.10	0.15
N	572,169	572,169	572,169	466,378
Panel D: Inventor Age, Extended Sample, 1985-2004				
inventor age	-0.201	-0.327	-2.359	-0.046
	(0.010)	(0.036)	(0.134)	(0.011)
R^2	0.27	0.15	0.19	0.25
N	1,855,887	1,855,887	1,855,887	1,550,825
Panel E: CEO Age and Inventor Age, Unbalanced Firm Sample, 1992-2004				
inventor age	-0.233	-0.438	-2.876	-0.019
	(0.026)	(0.121)	(0.321)	(0.022)
CEO age	-0.119	-0.317	-1.218	0.028
	(0.036)	(0.126)	(0.388)	(0.022)
R^2	0.14	0.03	0.09	0.15
N	316,516	316,516	316,516	263,641

Notes: Patent-level panel regressions with annual observations. The dependent variables are innovation quality at the patent level; a dummy for the patent being above the 99th percentile of the citation distribution; dummy for the patent being above the 90th percentile of the citation distribution; and generality index at the patent level (the last three are multiplied by 100 to ease legibility). Robust standard errors clustered at the firm level are in parentheses. Panel A is for our unbalanced firm sample 1992-2002 and controls for log employment, log sales, log patents, a full set of firm fixed effects, and application year dummies, and the key right-and side variable is CEO age. Panel B is for our unbalanced firm sample 1992-2002 and controls for log employment, log sales, log patents, application year dummies, a full set of firm fixed effects, a full set of dummies for inventor team size, a full set of dummies for three-digit IPC technology class dummies, and a full set of dummies for the total number of patents of the inventor within the team with the highest number of patents, and the key right-and side variable is average inventor age. Panel C expands the sample of Panel B to 1985-2002 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 1992-2002 and adds CEO age to the specification of Panel B. See text and notes to Table 1 for variable definitions.

Table 9: Inventor Age and CEO Age,
Unbalanced Firm Sample, 1992-2004

	Inventor age (1)	Inventor age (2)
CEO age	0.014	0.013
	(0.006)	(0.002)
R^2	0.11	0.13
N	316,516	316,516

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

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

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
CEO age	-0.180	-0.216	-0.087	-0.044
	(0.027)	(0.027)	(0.017)	(0.016)
log sales	1.465	2.081	0.285	1.201
	(0.449)	(0.611)	(0.272)	(0.328)
log patent	-0.394	-0.072	0.391	-0.020
	(0.193)	(0.257)	(0.136)	(0.151)
CEO age $\times$ log patent	-0.005	-0.071	-0.016	-0.037
	(0.014)	(0.021)	(0.011)	(0.011)
CEO age $\times$ log sales	0.024	0.079	0.009	0.044
	(0.017)	(0.021)	(0.012)	(0.011)
R^2	0.67	0.55	0.31	0.77
N	7,111	7,111	5,803	6,232

Notes: Weighted firm-level panel regressions with annual observations for the unbalanced firm panel, 1992-2002, with number of patents (in that year) as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality (the last three are multiplied by 100 to ease legibility). 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.

Table 11: Baseline Cross-Country Regressions

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
Panel A: Average Manager Age				
manager age	-0.484	-0.960	-0.225	-0.278
	(0.225)	(0.221)	(0.058)	(0.056)
log income per capita	-0.491	-0.702	-0.136	0.211
	(1.153)	(1.066)	(0.291)	(0.468)
secondary years of schooling	-1.000	-1.359	-0.291	-0.231
	(1.481)	(1.462)	(0.396)	(0.341)
log patent	2.232	2.331	0.591	1.072
	(0.706)	(0.695)	(0.193)	(0.222)
R^2	0.74	0.82	0.80	0.80
N	37	37	37	37
Panel B: Individualism				
individualism	4.965	9.929	2.369	3.420
	(2.461)	(2.393)	(0.640)	(0.487)
log income per capita	-1.233	-2.130	-0.472	-0.252
	(1.195)	(1.270)	(0.334)	(0.373)
secondary years of schooling	-0.467	-0.317	-0.056	-0.051
	(1.229)	(1.174)	(0.323)	(0.227)
log patents	1.622	1.125	0.308	0.725
	(0.490)	(0.472)	(0.129)	(0.164)
R^2	0.73	0.81	0.79	0.83
N	50	50	50	50
Panel C: Uncertainty Avoidance				
uncertainty avoidance	-8.354	-13.528	-3.174	-4.242
	(2.946)	(2.715)	(0.722)	(0.798)
log income per capita	-0.408	-0.657	-0.124	0.232
	(0.957)	(0.600)	(0.177)	(0.558)
secondary years of schooling	-0.745	-0.346	-0.054	0.008
	(1.149)	(1.108)	(0.307)	(0.208)
log patent	1.708	1.257	0.339	0.765
	(0.439)	(0.424)	(0.125)	(0.189)
R^2	0.80	0.86	0.84	0.84
N	50	50	50	50

Notes: Weighted cross-country regressions with total number of patents as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality (the last three are multiplied by 100 to ease legibility). See text and notes to Table 1 for variable definitions. Each country observation is the sample average between 1995-2000 as described in the text and the notes to Table 1. Robust standard errors are in parentheses.

Table 12: Cross-Country Regressions (Alternative Measures)

	Innovation Quality (5 years)	Superstar Fraction (Best Patent)	Tail Innovation (90/50)	Originality
	Panel A: Average Manager Age			
manager age	-0.203	-0.005	-1.002	-0.713
	(0.092)	(0.004)	(0.372)	(0.083)
R^2	0.75	0.80	0.70	0.88
N	37	37	37	37
	Panel B: Individualism			
individualism	2.039	0.052	9.966	8.015
	(1.009)	(0.045)	(4.028)	(0.653)
R^2	0.74	0.80	0.68	0.91
N	50	50	50	50
	Panel C: Uncertainty Avoidance			
uncertainty avoidance	-3.461	-0.106	-15.964	-9.084
	(1.215)	(0.057)	(4.689)	(1.336)
R^2	0.81	0.83	0.78	0.87
N	50	50	50	50

Notes: Weighted cross-country regressions with total number of patents as weights. The dependent variables are alternative measures of innovation quality (computed over the next five years), superstar fraction (with superstars defined according to the best patent), tail innovation (with fraction of patents above the 90th percentile of the citation distribution in the numerator), and the originality index (the last three are multiplied by 100 to ease legibility). Each regression also controls for log income per capita, average years of secondary schooling, and log total patents. See text and notes to Table 1 for variable definitions. Each country observation is the sample average between 1995-2000 as described in the text and the notes to Table 1. Robust standard errors are in parentheses.

Table 13: Cross-Country Regressions (Controlling for R&D Intensity)

	Innovation Quality	Superstar Fraction	Tail Innovation	Generality
	Panel A: Average Manager Age			
manager age	-0.636	-1.096	-0.257	-0.622
	(0.255)	(0.253)	(0.066)	(0.105)
R^2	0.76	0.83	0.81	0.91
N	33	33	33	33
	Panel B: Individualism			
individualism	8.245	13.786	3.291	2.932
	(2.821)	(2.602)	(0.725)	(0.778)
R^2	0.78	0.85	0.83	0.83
N	44	44	44	44
	Panel C: Uncertainty Avoidance			
uncertainty avoidance	-9.589	-14.173	-3.305	-3.452
	(2.747)	(2.753)	(0.754)	(0.915)
R^2	0.82	0.86	0.83	0.85
N	44	44	44	44

Notes: Weighted cross-country regressions with total number of patents as weights. The dependent variables are innovation quality, superstar fraction, tail innovation, and generality (the last three are multiplied by 100 to ease legibility). Each regression also controls for log income per capita, average years of secondary schooling, log total patents, and R&D intensity defined as total R&D expenditure divided by GDP. See text and notes to Table 1 for variable definitions. Each country observation is the sample average between 1995-2000 as described in the text and the notes to Table 1. Robust standard errors are in parentheses.

Table A1: Average Annual Patent Counts by Country, 1995-2000

Country	*Abbreviation*	*Patent Count*	*Country*	*Abbreviation*	*Patent Count*
Argentina	AR	9.2	India	IN	90.3
Austria	AT	365.0	Italy	IT	1439.8
Australia	AU	744.0	Japan	JP	33954.8
Belgium	BE	522.8	South Korea	KR	3581.5
Bulgaria	BG	3.8	Luxemburg	LU	62.8
Brazil	BR	69.7	Malta	MT	2.0
Canada	CA	2433.2	Mexico	MX	59.2
Switzerland	CH	1588.7	Malaysia	MY	14.5
Chile	CL	8.8	Netherlands	NL	1236.7
China	CN	109.5	Norway	NO	239.2
Colombia	CO	2.0	New Zealand	NZ	104.7
Czech Republic	CZ	17.7	Poland	PL	10.0
Germany	DE	9257.0	Portugal	PT	8.7
Denmark	DK	448.5	Romania	RO	2.7
Spain	ES	193.8	Russia	RU	88.2
Finland	FI	910.3	Saudi Arabia	SA	18.2
France	FR	3877.5	Sweden	SE	1691.3
Great Britain	GB	2869.5	Singapore	SG	191.2
Greece	GR	15.7	Slovenia	SI	13.7
Hong Kong	HK	171.8	Slovakia	SK	4.0
Croatia	HR	7.7	Thailand	TH	10.7
Hungary	HU	33.3	Turkey	TR	5.3
Indonesia	ID	3.0	United States	US	93722.5
Ireland	IE	111.3	Venezuela	VE	24.3
Israel	IL	580.7	South Africa	ZA	88.7

Notes: This table shows the average annual patent counts between 1995-2000, registered at the USPTO from that country.

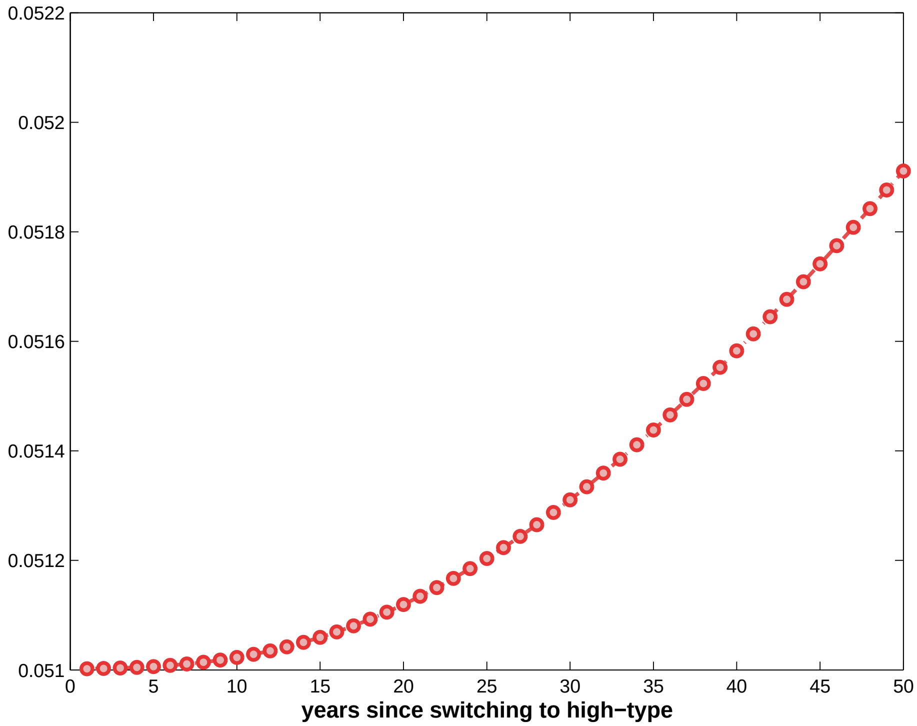


Figure 1: Evolution of creative innovations for high-type firms.

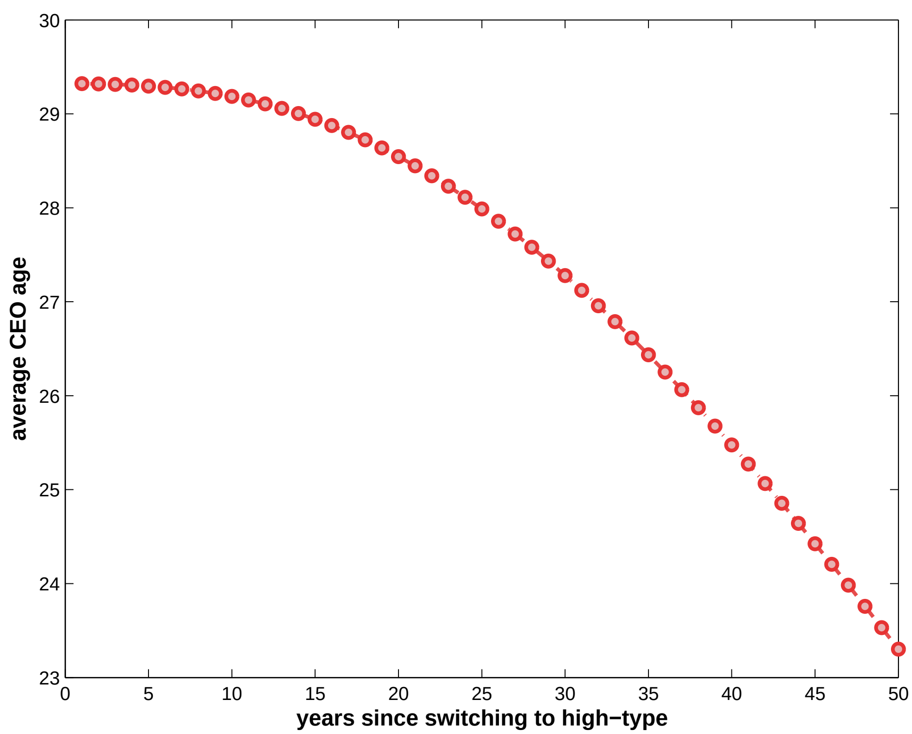


Figure 2: Evolution of CEO age for high-type firms.