The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?*

Benjamin F. Jones

14 January 2002

Abstract

This paper investigates, theoretically and empirically, a possibly fundamental aspect of technological progress. If knowledge accumulates as technology progresses, then successive generations of innovators may face an increasing educational burden. Innovators can compensate in their education by choosing narrower expertise, but narrowing expertise will reduce their individual capacities, with implications for the organization of innovative activity – a greater reliance on teamwork – and negative implications for growth. I develop a formal model to examine the growth implications of this "knowledge burden mechanism" and generate several testable predictions for innovators. The model predicts that educational attainment will rise over time and defines conditions under which specialization and hence teamwork will increase. In cross-section, the model predicts that specialization and teamwork will be greater in deeper areas of knowledge while, surprisingly, educational attainment will not vary across fields. I test these predictions using a micro-data set of individual inventors and find evidence consistent with each of these predictions. The model can thus provide a parsimonious explanation for a range of empirical patterns of inventive activity. The knowledge burden mechanism suggests that the nature of innovation is changing, with negative implications for long-run economic growth.

*I wish to thank Pol Antras, Andrei Bremzen, Esther Duflo, Glenn Ellison, Amy Finkelstein, Simon Johnson, Ben Olken, Michael Piore, and participants at various lunches and seminars for helpful comments. I am especially grateful to Daron Acemoglu, Abhijit Banerjee, and Sendhil Mullainathan for their advice and Trevor Hallstein for research assistance. The support of the Social Science Research Council's Program in Applied Economics, with funding provided by the John D. and Catherine T. MacArthur Foundation, is gratefully acknowledged.

1 Introduction

The importance of technological progress to growth is well accepted, yet the process of technological progress is not well understood. The recent growth literature, starting with the seminal contribution of Romer (1990), has made significant strides by considering technological advances as the output of rational agents operating in an explicit R&D sector. This approach seems realistic and succeeds in producing an endogenous, policy-variant description of the evolution of productivity. At the same time, this literature has highlighted the critical role of assumptions regarding the "knowledge production function", which defines how effort in R&D is mapped into productivity enhancements. Is growth a steady process where a given amount of research effort can produce constant productivity growth, or is innovation "getting harder" in the sense that a given amount of research effort will have a declining impact on growth over time?

The answer to this question has important implications, not just for the nature of technological progress, but also for the long-run growth potential of the world economy. If innovation is getting harder, a view associated with Jones (1995a, 1995b), Kortum (1997), and Segerstrom (1998), then steady growth in productivity is seen to rely on an ever-increasing level of innovative effort. Put in stark terms, if the world economy cannot indefinitely grow its research effort, then productivity growth will eventually cease. Jones, Kortum, and Segerstrom cite a range of evidence to support such a view, which I review briefly in Section 2.

In this paper I investigate, both theoretically and empirically, a mechanism through which technological progress may become harder with time. I start with the observation that technology in an economy is associated with a large body of knowledge. Innovators are not born at the frontier of knowledge; instead, they must undertake education – if one is to stand on the shoulders of giants, one must first climb up their backs. If technological progress leads to an accumulation of knowledge, then the educational burden on successive generations of innovators will increase. Innovators may compensate by choosing narrower expertise: a "death of the Renaissance Man" effect. This narrowing of expertise reduces the capabilities of individual innovators, which in turn has implications for the organization of innovative activity and growth in the economy.

To help motivate this mechanism, consider the invention of the microprocessor. As described by Malone, the invention started with the inspiration of a researcher named Ted Hoff:

> [Hoff] had been working with a DEC computer doing circuit design and had been impressed by how the computer could do such complex tasks...this, he thought, could be his model for a new type of circuitry.

Hoff, at Intel, teamed up with Stan Mazor and Masatoshi Shima. Together they developed Hoff's idea:

> Hoff's greatest contribution was the logic chip and the design of the chip set's architecture; Shima's was in the controller chip and chip set's logic.

To implement their design, however, they were forced to turn to another specialist:

> Hoff and Mazor didn't really know how to translate this architecture into a working chip design. And with that, they didn't know whether there were any flaws in their architecture. The project began to lag.
>
> In fact, probably only one person in the world did know how to do the next step. That was Federico Faggin... (Malone, 1995)

The microprocessor was one person's inspiration, but four people's invention. It is the story of researchers with circumscribed abilities, working in a team, and it helps motivate the model of innovation and growth explored in this paper.

This paper proceeds as follows. In Section 2, I review the existing debate in the growth literature over whether innovation is getting harder. Theories in the literature for why innovation may or may not be getting harder tend to be suggestive and, where mechanisms are formulated more explicitly, difficult to test. As a result, the empirical work has tended to rely on reduced-form predictions that are tested with data aggregates. The reliance on data aggregates has in turn allowed much room for debate. The knowledge burden model presented in this paper will make predictions that are consistent with the evidence cited in this literature and will also incorporate leading ideas from this growth debate. In addition, the knowledge burden mechanism suggests a number of specific, further tests that do not rely on data aggregates and are not easily explained by existing theory.

The model is presented in Section 3. Innovators make costly education decisions in an economy that may, over time: (i) produce more or less knowledge that would-be innovators need to learn, (ii) produce rising or falling technological opportunities; and (iii) grow its population. Education is valuable to innovators and its value is complementary to income possibilities in the innovative sector. Along the steady-state growth path, these income possibilities expand – due to increasing market size if nothing else – so that new cohorts will seek more education over time. If the knowledge burden mechanism is sufficiently strong then successive cohorts of innovators, despite their greater educational achievement, will compensate by choosing a narrower range of expertise, with negative implications for their individual capabilities. The model thus suggests a growing burden of knowledge as an independent channel through which innovation may become more difficult with time. It raises the bar on other mechanisms in the evolution of innovators' productivity – asking more of optimistic stories if we wish to preserve the possibility of steady-state growth without relying on exponentially increasing effort. Furthermore, and perhaps more importantly, the model makes several specific predictions about the behavior of individual innovators. In time series, the model predicts that educational attainment will be rising and defines conditions under which specialization and hence teamwork will increase. In cross section, the model predicts that specialization and the propensity to form teams will be greater in fields where knowledge is deeper. At the same time, income arbitrage in the model ensures that educational attainment will not vary across technological fields, regardless of variation in the depth of knowledge or innovative opportunities.

Section 4 explores the predictions of the model empirically. Using a rich patent data set (Hall et al. 2001) together with the results of a new data collection exercise to determine the ages of 55,000 inventors, I am able to develop detailed patent histories for individuals. I find that the age at first innovation is trending upwards at 0.6 years per decade, specialization is increasing at 6% per decade, and U.S. team size is increasing at 17% per decade. In cross-section, I find support for the model's perhaps less obvious prediction that educational attainment will be similar across fields. At the same time, team size and the specialization measure vary substantially across fields, and, as predicted, vary in a supportive manner when related to a direct measure of the amount of prior art underlying each patent. The knowledge burden mechanism thus serves as a single, parsimonious explanation for this collection of new facts, as well as for the existing facts to be presented in Section 2, with negative implications for growth.

Section 5 concludes.

2 Existing Evidence and Debate

Jones (1995a, 1995b), Kortum (1997), and Segerstrom (1998) cite several trends to support the view that steady-state growth is relying on growing research effort:

1. R&D expenditures and R&D employment are rising dramatically in the U.S., Japan, Germany, and France, while TFP growth in these countries is flat (Jones 1995a).

2. The ratio of patent counts to R&D employment is falling over time in all countries (Evenson 1984) and since 1870 in the U.S. (Machlup 1962). The ratio of patent counts to R&D expenditures is also falling dramatically over time across U.S. manufacturing industries (Kortum 1993).[1]

Figure 2.1 shows that TFP growth in the U.S. had been flat despite large increases in R&D employment and R&D expenditures. Figure 2.2 shows the recent decline in U.S. patent grants per U.S. R&D worker.[2]

These facts pose a serious challenge to growth models in which a fixed amount of research effort produces steady growth (e.g. Romer 1990, Grossman & Helpman 1991, Aghion & Howitt 1992, 1998).[3] The challenge for such models is how to simultaneously explain constant growth rates given the apparent increase in research effort – the so-called "scale effects" problem. Whether this challenge has been or will be met is an open question; what is clear is that the reliance on data aggregates in these arguments has left much room for debate. First, criticisms of the data can certainly be made: our productivity measures may be poor, as may our aggregate measures of R&D effort, and patent counts may be a

[1] Kortum and Segerstrom also cite case studies of the pharmaceutical industry (Henderson & Cockburn 1996), the textiles and chemical industries (Baily & Chakrabarti 1985) and the microprocessor industry (Malone 1995), which describe a sense among researchers in these industries that innovation is becoming more difficult.

[2] To support the use of patent counts as a measure of inventive output, Kortum and Segerstrom further cite Mansfield (1986), who in a survey of the R&D departments of a random sample of 100 firms found that they reported no decreasing propensity to patent inventions over time.

[3] Kremer (1993) provides further evidence to challenge these models. Using data on population over the last one million years and a Malthusian model where population is limited by technology, Kremer estimates that the research productivity of individuals has increased at only two-fifths the rate necessary to provide steady growth without increasing effort.

poor measure of inventiveness (though see footnote 2). Second, the aggregated data leaves broad room for interpretation; for example, Young (1998) and others have succeeded in explaining observation #1 using "expanding product space" models that produce steady growth both with and without an increase in effort.[4]

Progress in this debate will be aided by defining specific, testable mechanisms through which the productivity of innovators may rise or fall as the economy develops. Heretofore, the theoretical arguments in the literature have tended to be suggestive. Some authors point to "fishing out" hypotheses – where big ideas are progressively harder to come by. Other authors point to the possibility of "positive intertemporal spillovers" in knowledge production – where the introduction of faster computers, the Internet, and key ideas like calculus and Newtonian physics may enhance future innovators' productivity. Where such mechanisms have been given rigorous microfoundations, the mechanisms appear difficult to test.[5]

This paper proposes a growing burden of knowledge as a specific mechanism through which innovation may become more difficult over time. By focusing on innovators as the unit of analysis, the model produces several implications for the behavior of individual innovators, allowing tests of the theory which do not rely on data aggregates.

3 The Model

The over-arching theme of this model is the emphasis on innovators. I analyze a simple structure with two sectors: a production sector where competitive firms produce a homogenous output good and an innovation sector where innovators produce productivity-enhancing ideas. Workers in the production sector earn a competitive wage while innovators earn

[4] Such models succeed by (1) limiting the impact of research to specific product lines, and (2) assuming that the number of product lines increases in exact proportion with the population. These product-space models thus neutralize the growth effects of increasing population (and consequent increases in the scale of research effort); therefore, these models can explain observation #1. At the same time, they are precariously balanced (see Jones 1999) and provide no explanation for observation #2.

[5] Two rigorous theories for the evolution of the knowledge production function should be noted. Kortum (1997) models innovations as draws from a random distribution and defines generally how this distribution must evolve to absorb the scale effect of population. While useful, and consistent with the aggregate facts presented above, his model does not explain why the distribution of ideas should evolve in any particular way. Weitzman (1998) presents explicit microfoundations for the knowledge production function, arguing that ideas are combinatoric in nature: the production of new ideas leads to combinatoric (i.e. greater than exponential) growth in the number of further ideas to try. The limitation in growth becomes the rate at which innovators can examine the possibilities. Weitzman's ultimate result that human capacities are the limiting factor in growth is similar to the themes of this paper.

income by licensing their ideas to firms in the production sector. I abstract from physical capital in the model and focus on the role of human capital in the innovation sector. Innovators must undertake a costly human capital investment to bring themselves to the knowledge frontier where they become able to innovate. Innovators face a tradeoff between the costs of seeking more education and the benefits of achieving a broader degree of expertise. This tradeoff will be balanced differently by different cohorts as the amount of knowledge in the economy evolves.

Section 3.1 describes the production sector and Section 3.2 defines individuals' life-cycles and preferences. Sections 3.3 and 3.4 focus on innovators. The first describes the knowledge space, the innovator's choice of specialty, and the cost of education. The second considers the process of innovation, the value of innovations, and the evolution of innovators' productivity. Section 3.5 defines individuals' equilibrium choices. Section 3.6 analyzes steady-state growth, relating the predictions of this model back to the discussion in the introduction. Section 3.7 examines the time-series predictions of the model. Section 3.8 extends the model to investigate its predictions across technological areas at a point in time. The predictions of Sections 3.7 and 3.8 are the foundation for the empirical analysis in Section 4.

3.1 The Production Sector

Competitive firms in the production sector produce a homogenous output good. A firm j hires an amount of labor, $l_j(t)$, to produce output,

$$y_j(t) = X_j(t) l_j(t) \tag{1}$$

which it sells at the numeraire price, $p_y(t) = 1$. $X_j(t) \leq X(t)$ is the productivity level of firm j, where $X(t)$ is the leading edge of productivity in the economy, which can be achieved by any firm with access to the entire set of productive ideas that have been produced by the innovative sector.

The firm pays workers a wage, $w(t)$, and makes royalty payments per worker of $r(t)$ on any patented technologies it employs. While patent protection lasts, the monopolist innovator will charge a firm a fee, per period, equivalent to all the extra output the firm can produce with the innovation, and the firm will be just willing to pay this fee. Therefore $X_j(t) = X(t)\ \forall j$, and the total output in the economy is:

$$Y(t) = X(t)L_Y(t) \tag{2}$$

The revenues of these competitive firms are dispensed entirely in wage and royalty payments, $X(t)l_j(t) = w(t)l_j(t) + r(t)l_j(t)$. The competitive wage paid to a production worker is therefore:

$$w(t) = X(t) - r(t) \tag{3}$$

3.2 Workers and Preferences

There is a continuum of workers of measure $L(t)$ in the economy at time t. This population grows at rate g_L. Individuals face a constant hazard rate ϕ of death. The constant hazard rate model has well-known properties: the probability of surviving to time t given birth at time τ is $e^{-\phi(t-\tau)}$, and a worker's life expectancy at any point in time is $1/\phi$.

Individuals are risk-neutral and share a common intertemporal utility function,[6]

$$U(\tau) = \int_{\tau}^{\infty} c(t)e^{-\phi(t-\tau)}dt \tag{4}$$

Each individual faces a dynamic budget constraint, $da(t)/dt = \phi a(t) + f(t) - c(t)$, where $a(t)$ is her assets, $f(t)$ is her flow of non-interest income, and $c(t)$ is her consumption in period t. Note that, for simplicity, I assume that the hazard rate of death serves as both the rate of time preference and the interest rate in the economy. Individuals are therefore indifferent to the timing of their consumption; they are also indifferent to the riskiness of their income stream.

The choice problem of interest in the model is that of career. I assume that individuals are born without assets and supply a unit of labor inelastically at all points over their lifetime. From the standard intertemporal budget constraint, the individual's utility is equivalent to the present value of her expected lifetime non-interest income. At birth, an individual decides whether to become a wage worker or an innovator. Wage workers require no education and their expected utility is simply the discounted flow of the wage payments they receive:

[6] For simplicity of exposition, I will specify the incomes and expenditures in the model in terms of a unit-mass of individuals.

$$U^{wage}(\tau) = \int_{\tau}^{\infty} w(t) e^{-\phi(t-\tau)} dt \tag{5}$$

If an individual i chooses to be an innovator instead, then she must further choose a specific field of expertise and pay an immediate fixed cost of education, E, to bring herself to the frontier of knowledge in that area. Having paid this cost, the innovator earns an expected flow of income, v, by licensing any innovations she produces to firms in the production sector. In the model, both v and E are specific to the choice of expertise made by an individual (i). Income and the educational cost will also depend on the time of birth (τ), and income flows will further depend on the current state of the economy (t). The expected lifetime utility of an innovator is written generally as,

$$U_i^{R\&D}(\tau) = \int_{\tau}^{\infty} v_i(\tau, t) e^{-\phi(t-\tau)} dt - E_i(\tau) \tag{6}$$

The structure of the innovator's educational choice and the functional forms of v and E are the subject of the next two subsections.

3.3 Knowledge and Education

A type of knowledge is defined by its position, s, on the unit circle. For example, one segment of the circle might represent electronics, another biochemistry, another economics. At a point in time, the amount of knowledge at each point on the circle is assumed to be the same.[7] I define this quantity as $D(t)$.

The prospective innovator chooses an area of expertise: a point, s_i, on the circle and a certain distance, $b_i \in [0, 1]$, to its right. For an innovator born at time τ, the amount of knowledge the innovator acquires is the chosen breadth of expertise, b_i, multiplied by the prevailing depth of knowledge, $D(\tau)$. The educational cost of acquiring this information is:

$$E_i(\tau) = (b_i D(\tau))^{\varepsilon} \tag{7}$$

where $\varepsilon > 0$, which says only that learning more requires a greater amount of education. I make no a priori assumption about whether education costs are convex or concave in the amount of information the innovator learns.

[7] I will partly relax this assumption when I consider a cross-sectional variation of the model in Section 3.8.

With the assumption that the depth of knowledge is evenly arrayed around a unit circle, the total depth of knowledge at a point in time is $D(t)$. In general, the depth of knowledge will change as innovators produce new ideas. However, while these new ideas serve to increase the productivity in the economy, $X(t)$, they may or may not increase $D(t)$. I write,

$$D(t) = (X(t))^{\delta} \tag{8}$$

with no assumption regarding the sign of δ. It may be natural to assume that the production of new ideas in the R&D sector leads to an increase in $D(t)$. However, we might also imagine that new ideas either replace old ideas or simplify ideas so that $D(t)$ may actually fall as productivity rises. This latter interpretation is consistent with the concept of revolutionary "paradigm shifts", which Thomas Kuhn has suggested as the appropriate model of scientific progress (Kuhn 1962).

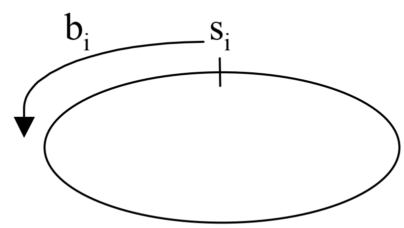

Figure 3.1: The circle of knowledge

3.4 Innovation

Once educated, innovators begin to receive innovative ideas. Ideas arrive randomly, with hazard rate λ for a unit-mass of individuals. When an idea arrives, it comes with two further properties. The first is the random breadth of expertise, k, required to implement the idea. The second is the size of the idea, which adds to TFP by an amount γ.[8]

The required expertise, k, may be greater or less than the inspired innovator's own expertise, b_i. For example, a statistician might conceive of a new statistical method and be able to implement the idea solely on the basis of her own expertise. An engineer might have an idea for a new space shuttle for NASA, but the implementation requires much broader expertise than the engineer himself possesses. The breadth of the idea $k \in [0, 1]$ is drawn

[8] One can imagine more generally that the size of ideas is random, where γ is the mean size; this interpretation has no effect on the model.

from a smooth distribution function F. It is measured as a distance to the right from an individual's location s_i, so that the implementation of the idea requires expertise over the segment of the circle $[s_i, s_i + k]$. Therefore, with probability $F(b_i)$ the innovator is able implement the idea alone, and with probability $1 - F(b_i)$ the innovator needs at least one partner. That is, I allow for the formation of teams.

I assume that the innovator with the idea acts as a monopolist vis-a-vis potential teammates so that, by Bertrand reasoning, the inspired innovator receives all profits from the project. I further assume that once an idea arrives it can be implemented instantaneously and without any expenditure (in particular, team formation is costless). Therefore, (i) all projects are profitable, (ii) the inspired, monopolist innovator will receive the entire royalty stream from the project as personal income, and (iii) any necessary teammates will be just willing to help without compensation.

The only possible obstacle to implementation is an absence of required expertise. Anticipating the equilibrium of this model, innovators' collective expertise will cover the entire circle of knowledge, so that all ideas are feasible and therefore all ideas will in fact be implemented. To avoid burdensome notation in the text, I will write the rest of the model assuming this result. The Appendix considers the general case and establishes this result as part of any (subgame perfect) equilibrium.

I make two further assumptions regarding team formation. First, the inspired innovator will choose team members from her own cohort if possible. Second, the innovator assembles the minimum number of people necessary to cover the breadth of expertise, k, required to implement the idea. These last two assumptions are innocuous and are made to permit explicit analysis of average team size, which is explored in Sections 3.7 and 3.8.

Given that an idea increases TFP by an amount γ, it can be licensed for use by L_Y workers, and patent protection lasts for z years, the lump-sum value of the patent is:[9]

$$V = \gamma \int_t^{t+z} L_Y(\widetilde{t}) d\widetilde{t} \tag{9}$$

[9] This expression is written assuming the innovator has access to a competitive financial market which will pay the innovator the lump-sum value of the patent (or an equivalent annuity) in exchange for the patent rights. If no such market were available, the value of the patent to the innovator would need to reflect the possibility that the innovator dies before the patent rights expire, in which case $V = \varphi \int_t^{t+z} L_Y(\widetilde{t}) e^{-\phi(\widetilde{t}-t)} d\widetilde{t}$. If the latter route is taken, we need to assert additionally that any remaining patent rights are assigned upon the innovator's death in a way that has no asymmetric effect on the incomes of the rest of the population. This variation will have no impact on the main results of the model.

Along the balanced growth path, the fraction of production workers, $L_Y(\widetilde{t})/L(\widetilde{t})$, will be constant. $L_Y(\widetilde{t})$ thus grows at g_L and we can integrate (9) to find: $V = \gamma C L_Y$, where $C = (e^{g_L z} - 1)/g_L$, and L_Y is the mass of production workers at the time of the innovation.

The expected flow of income to an innovator is $v = \lambda V$, the probability of having an idea at a point in time times the income the idea generates. Using the definition of V, we can write $v = \lambda\gamma C L_Y$. The expected flow of income can therefore equivalently be understood as the expected rate at which the innovator adds to TFP, $\lambda\gamma$, times the market size for the innovation, CL_Y. When considering the time lag between an innovator's innovations (Section 3.7) and associated empirical analysis (Section 4), it will be useful to consider λ and γ separately. However, for the main analysis of the model, which considers steady-state growth, I wish to emphasize that the combination of these parameters is the important primitive. I will therefore also define $\theta = \lambda\gamma$ as a summary measure of innovator productivity.

The parameters λ and γ will in general differ across individuals (i), across cohorts (τ), and across time (t). Specifically, I assume that λ and γ, and hence θ and v, will depend on three things: (1) the vintage of knowledge the innovator learns; (2) the current degree of competition in the innovator's specialty; and (3) the innovator's breadth of expertise. In particular, I write,

$$\lambda_i(\tau, t) = X(\tau)^{\chi_\lambda} L(t, s_i)^{-\sigma} b_i^{\beta_\lambda} \tag{10}$$

$$\gamma_i(\tau, t) = X(\tau)^{\chi_\gamma} b_i^{\beta_\gamma} \tag{11}$$

and therefore

$$\theta_i(\tau, t) = X(\tau)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} \tag{12}$$

where $\chi = \chi_\lambda + \chi_\gamma$ and $\beta = \beta_\lambda + \beta_\gamma$. $X(\tau)$ is the productivity level in the economy at the innovator's birth, $L(t, s_i)$ is the mass of individuals at time t who share the innovator's specialty, and b_i is the innovator's breadth of expertise.

These reduced-form specifications capture several key ideas. The parameter $\chi = \chi_\lambda + \chi_\gamma$ represents the impact of the state of technology on an innovator's productivity. It incorporates the standard ideas in the literature which were discussed in Section 2: "fishing-out" hypotheses whereby innovators' productivity falls as the state of knowledge advances

($\chi < 0$), and "positive intertemporal spillovers" whereby an improving state of knowledge makes innovators more productive ($\chi > 0$).[10,11]

The parameter σ represents the impact of crowding on the frequency of an innovator's ideas. I assume $\sigma > 0$, following standard arguments where innovators partly duplicate each other's work. A greater density of workers in the same specialty increases competition, reducing the rate at which a specific individual produces a novel idea.

The final parameter, $\beta = \beta_\lambda + \beta_\gamma$, represents the impact of the breadth of expertise. A specification with $\beta > 0$ suggests simply that greater human capital increases one's productivity. The specific reason I embrace, for the purposes of this model, is that individuals with broader expertise access a larger set of available knowledge – facts, theories, methods – on which to build innovations. This will increase their innovative abilities, along the lines of Weitzman (1998), making them more productive.[12]

With the definitions (10) and (11), I can now explicitly define an innovator's expected flow of income,

$$v_i(\tau, t) = X(\tau)^\chi L(t, s_i)^{-\sigma} b_i^\beta C L_Y(t) \tag{13}$$

3.5 Equilibrium Choices

The choice facing each individual is that of career, which is a one-shot decision made at birth. Players have infinitesimal mass so that the actions of any specific individual do not

[10] Note that I am using the state variable $X(\tau)$ to represent the effect of both technology and the state of knowledge on an innovator's capabilities. We could introduce a second state variable, $A(\tau)$, to represent the state of knowledge and add a separate channel through which the quality of existing ideas influences an innovator's abilities. Since the state of knowledge in standard growth models is assumed to be deterministically related to the technology level in the economy, adding a separate channel to differentiate between "ideas" and "technology" will add little insight. When I discuss cross-sectional predictions in Section 3.8, where it will be useful to think of different knowledge levels across technological areas, I will introduce a richer specification.

[11] By writing (12) I assume that only the vintage of knowledge and productivity at birth matter. A more general specification would allow innovators' productivity to improve to some degree as technology or knowledge in the economy improve over their lifetime, but such a specification adds no important intuition to the model, so it is left out for simplicity.

[12] There are many other mechanisms through which broader expertise would enhance an innovator's income. First, a more broadly expert innovator may better evaluate the expected impact and feasibility of her ideas. She will better select toward high value, successful lines of inquiry, and therefore achieve greater returns. Second, if assembling teams is costly, innovators will be unwilling to form large teams. More broadly expert innovators can rely less on large teams for the implementation of their ideas, making their ideas less costly to implement. Third, if income is shared across team members, then broader expertise, which reduces the necessary team size, will bring one a greater share of project income. These last two effects will lead more narrowly expert innovators to abandon a greater portion of their broad ideas.

influence the income of others. At the same time, an individual's income will depend on the collective decisions of other players. To rule out possible pathological multiple equilibria, I assume that only strictly positive masses of workers are observable to players, so that strategies cannot be conditioned on the actions of specific individuals.

Define the set of individuals born at time τ as $l(\tau)$, of which a subset $l_Y(\tau)$ choose the production sector and a subset $l_R(\tau)$ choose the innovation sector instead. Those who choose the innovation sector must additionally choose an area of expertise (s, b). In equilibrium, we require two conditions for each subgame τ:

$$U_i^{R\&D}(s_i, b_i) \geq U_i^{R\&D}(s, b) \quad \forall s, b \ \ \forall i \in l_R(\tau) \tag{14}$$

$$U_i^{R\&D}(s_i, b_i) = U^{wage} \quad \forall i \in l_R(\tau) \ \ \forall j \in l_Y(\tau) \tag{15}$$

The first condition states that no innovator can deviate to any other choice (s, b) and be better off. The second condition rules out income arbitrage possibilities between the R&D and production sectors.[13] With the definitions of the model in Sections 3.1 through 3.4, we can now define the expected income from various choices and hence, with conditions (14) and (15), the equilibrium outcome.

3.5.1 Production workers

Production workers receive a competitive wage $w(t) = X(t) - r(t)$, where $X(t)$ is the leading edge of productivity in the economy and $r(t)$ is the royalty payments the firm makes per worker to access the latest technologies. To define the flow of royalty payments, note that the expected creation of royalties in any interval dt is dX. Since patents are protected for z years, the flow of royalty payments $r(t)$ is then:

$$\int_{t-z}^{t} dX$$

which is simply $X(t) - X(t-z)$. The wage worker's expected flow of income at any time t is then $w(t) = X(t) - r(t) = X(t-z)$. In other words, the wage earned by a production worker is that portion of productivity which is not patent-protected, which is just the productivity level of the economy z years previously.

[13] Condition (15) is a reduced form of two separate conditions: (i) $U_i^{R\&D}(s_i, b_i) \geq U_i^{wage}$ $\forall i \in l_R(\tau)$; (ii) $U_j^{wage} \geq U_j^{R\&D}(s, b)$ $\forall s, b$, $\forall j \in l_Y(\tau)$. Noting that $U_i^{wage} = U_j^{wage}$ and $U_i(s_i, b_i) = U_j(s_i, b_i)$ for any two individuals in the same cohort, these two conditions reduce to (15).

Along a balanced growth path the growth rate in productivity is a constant, g, in which case $X(t-z) = X(t)e^{-gz}$. Assuming that $g < \phi$, so that workers have finite expected income, we integrate (5) to find:

$$U^{wage}(\tau) = \frac{X(\tau)}{\phi - g} e^{-gz} \tag{16}$$

3.5.2 Innovators

Using (6) and the equilibrium condition (14), the innovator's problem is:

$$\max_{s_i, b_i} \int_{\tau}^{\infty} v_i(\tau, t) e^{-\phi(t-\tau)} dt - E_i(\tau) \tag{17}$$

Or, with the definitions of $E_i(\tau)$ and $v_i(\tau, t)$ in (7) and (13),

$$\max_{s_i, b_i} \int_{\tau}^{\infty} X(\tau)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} C L_Y(t) e^{-\phi(t-\tau)} dt - (b_i D(\tau))^{\varepsilon} \tag{18}$$

First consider the choice of s_i. In the Appendix I prove that in any subgame perfect equilibrium of this game, $L(t, s) = L_R(t) \; \forall s$. That is, innovators evenly array themselves around the circle of knowledge. The intuition for this result is straightforward. Given that $\sigma > 0$, the integrand and hence expected income are strictly increasing as $L(t, s_i)$ falls. In consequence, the innovator seeks to avoid crowding and chooses a location in the circle of knowledge where the density of innovators is smallest. In equilibrium, no innovator will wish to deviate from her choice of s_i, in which case all innovators must array themselves evenly around the unit circle. The slight complexity in the proof is to consider – and rule out – the possibility that innovators array themselves so that there are "holes" in expertise around the circle and not all projects are feasible to implement. It should be intuitive that such holes cannot survive in equilibrium: for some innovator there will exist a small deviation into such a hole that will lead to congestion benefits without reducing the feasibility of her ideas. See the Appendix for the formal reasoning.

The innovator chooses b_i so that the marginal cost of education equals the marginal expected benefit to her income as an innovator. Differentiating (18) with respect to b_i produces the first order condition:

$$\frac{\beta}{b_i^*} C X(\tau)^{\chi} b_i^{*\beta} \int_{\tau}^{\infty} L_R(t)^{-\sigma} L_Y(t) e^{-\phi(t-\tau)} dt = \frac{\varepsilon}{b_i^*} (b_i^* D(\tau))^{\varepsilon} \tag{19}$$

One can readily verify, by looking at the second order condition, that this stationary point defines a (unique) maximum if and only if $\beta < \varepsilon$, which I will assume for the rest of the analysis.[14] Along the balanced growth path, $L_R(t)$ and $L_Y(t)$ will both grow at a constant rate equal to the growth rate in population, g_L. Using this property to evaluate the integral and rearranging the first order condition, we can characterize the equilibrium choice of b_i implicitly as,

$$b^*(\tau) = \frac{1}{D(\tau)} \left(\frac{v(\tau)}{\phi - (1-\sigma)g_L} \frac{\beta}{\varepsilon} \right)^{1/\varepsilon} \tag{20}$$

where I write $b^*(\tau)$ to acknowledge that innovators in the same cohort face the same maximization problem and therefore choose identical breadths of expertise.[15] Similarly, I write $v_i(\tau, \tau)$ as $v(\tau)$ to represent the expected flow of income to an innovator at the time of their birth. Given the optimal choice, $b^*(\tau)$, we can further define the equilibrium level of education and the expected utility for an innovator in cohort τ,

$$E^*(\tau) = \frac{v(\tau)}{\phi - (1-\sigma)g_L} \frac{\beta}{\varepsilon} \tag{21}$$

$$U^{R\&D*}(\tau) = \frac{v(\tau)}{\phi - (1-\sigma)g_L} \left(1 - \frac{\beta}{\varepsilon}\right) \tag{22}$$

Having defined the equilibrium choices of b and s, the remaining pieces of the equilibrium consider the labor allocation between the production and innovation sectors and the equilibrium determination of the growth rate. These three unknowns ($L_R(t)$, $L_Y(t)$, g) can be solved using three equations. The first is the arbitrage equation in expected lifetime income, equilibrium condition (15). The second is an accounting relationship for the allocation of labor, $L(t) = L_R(t) + L_Y(t)$. The third is the steady-state description of the growth rate, which is defined in the next section.

[14] If $\beta > \varepsilon$ then the first order condition defines a unique minimum, and if $\beta = \varepsilon$ it defines an inflection point. It is straightforward to show that in either case the innovator will choose the corner solution, $b_i^* = 1$. Such a corner solution can also emerge when $\beta < \varepsilon$ if the unique maximum described by (19) occurs where $b_i^* > 1$. These cases, where the innovator learns all available knowledge, are less interesting and will be left aside in further analysis.

[15] I leave the definition of b^* implicit because this formulation will be convenient for analyzing its growth path. We can write b^* explicitly by noting that $v(\tau) = X(\tau)^{\chi} L_R(\tau)^{-\sigma} b^{*\beta} C L_Y(\tau)$. Inserting this expression into (20) and rearranging shows that $b^*(\tau) = \left(\frac{1}{D(\tau)}\right)^{\varepsilon/(\varepsilon-\beta)} \left(\frac{C X(\tau)^{\chi} L_R(\tau)^{-\sigma} L_Y(\tau)}{\phi-(1-\sigma)g_L} \frac{\beta}{\varepsilon}\right)^{1/(\varepsilon-\beta)}$.

3.6 Steady-state Growth

Along the balanced growth path, the growth rate in per-capita income is equal to the growth rate in productivity, g. If there are $L_R(t)$ innovators active at a point in time and the average innovator raises productivity in the economy at a rate $\bar{\theta}(t)$, then productivity increases at rate $dX/dt = \bar{\theta}(t)L_R(t)$. The growth rate in the economy is then,

$$g = \frac{\bar{\theta}(t)L_R(t)}{X(t)} \tag{23}$$

This expression is mechanical and holds both inside and outside of steady-state. On the balanced growth path, where g is constant, we can take logs and differentiate with respect to time to see that $g = g_{\bar{\theta}} + g_L$. The growth rate is a function of the rate of population growth and the evolution of average innovator productivity. Using the equilibrium relationships we have previously derived, we will be able to express $g_{\bar{\theta}}$ as a function of g and various exogenously specified elasticities. The derivation is straightforward but uninformative and is presented in the Appendix. The steady-state growth rate derived there is,

$$g = \frac{1-\sigma}{1-\chi-\beta(\frac{1}{\varepsilon}-\delta)}g_L \tag{24}$$

which assumes that $\chi + \beta(\frac{1}{\varepsilon} - \delta) < 1$. This result, with its parametric condition, defines the growth rate as the outcome of several important forces. The parameter χ, as discussed above, represents standard ideas in the growth literature whereby the productivity of innovators may increase as they gain access to new technologies and new ideas ($\chi > 0$) or decrease if innovators are fishing out ideas ($\chi < 0$). The larger χ, the greater the growth rate, as is seen in (24). The parameter σ represents the degree to which increased research effort serves to duplicate existing effort. As σ approaches 1, research effort becomes increasingly congestive and the growth rate will tend toward zero.

The implications of an increasing burden of knowledge are contained in the term $\beta(\frac{1}{\varepsilon}-\delta)$. We can understand this term clearly by first considering the growth in the breadth of expertise. As shown in the Appendix in the derivation of (24),

$$g_{b^*} = \left(\frac{1}{\varepsilon}-\delta\right)g \tag{25}$$

This result implies that, on the balanced growth path, new cohorts of innovators become more specialized with time if and only if $\frac{1}{\varepsilon} - \delta < 0$, or equivalently iff $\varepsilon\delta > 1$. The first parameter, ε, is the elasticity of the cost of education with respect to the amount of knowledge an innovator learns. The second parameter, δ, is the elasticity of the depth of knowledge in the economy with respect to the level of technology. This specialization condition is intuitive: it says that people will specialize more with time if, in combination, education is sufficiently expensive and the depth of knowledge in the economy is rising at a sufficient rate. If this condition is satisfied, we will witness the "death of the Renaissance Man" along the growth path ($g_{b^*} < 0$). The impact of specialization on growth will be large or small depending on the value of β, the elasticity of innovators' productivity with respect to their breadth of expertise.

The growth rate given in (24) also shows that growth in per-capita income will depend on growth in the population. This is the standard Jones (1995b) style result discussed at the opening of this paper, where increasing effort is needed to produce steady-state growth. A growing population provides both the motive – increasing market size – and the means for innovative effort to grow at an exponential rate. The alternative, Romer (1990) style result, where growth can be sustained without an increase in effort, is obtained in the knife-edge case where $\chi + \beta(\frac{1}{\varepsilon} - \delta) = 1$. This parametric condition implies that the productivity of innovators increases in exact proportion with the productivity in the economy; hence, growth can be sustained with a fixed amount of effort (i.e. without population growth).[16] However, by the same token, if population is increasing, then growth rates will now explode (consider (24) with $g_L > 0$ as $\chi + \beta(\frac{1}{\varepsilon} - \delta) \to 1$). This is the usual "scale effects" problem, familiar from the growth literature and reviewed in Section 2. To produce positive but non-explosive growth rates both with and without population growth, we need to make the additional knife-edge assumption that $\sigma = 1$. This second knife-edge assumption absorbs the impact of increasing population by assuming that increased R&D effort is completely duplicative and has no impact on the growth rate.[17]

The greater the burden of knowledge in this model, i.e. the more negative $\beta(\frac{1}{\varepsilon} - \delta)$,

[16] With $g_L = 0$, we see from (23) that steady-state growth requires $g = g_{\bar{\theta}}$. As can be seen from (42) in the Appendix and (25), the productivity of innovators grows at a rate $g_{\bar{\theta}} = (\chi + \beta(\frac{1}{\varepsilon} - \delta))g$ when $g_L = 0$. Hence, in the absence of population growth, steady-state growth requires the knife-edge condition that $\chi + \beta(\frac{1}{\varepsilon} - \delta) = 1$.

[17] In consequence, the assumption that $\sigma = 1$ also eliminates the role of research subsidies in raising the growth rate. Models which use an expanding product space instead of congestion effects (e.g. Young 1998) to absorb the impact of population on the growth rate maintain the role of R&D subsidies in growth. Such models do not, however, avoid making two separate knife-edge assumptions. For a general discussion of the dual knife-edge properties of models with an increasing product space see Jones (1999).

then the larger χ must be to achieve the knife edge condition in which $\chi + \beta(\frac{1}{\varepsilon} - \delta) = 1$. Therefore, while not dispositive of other mechanisms, we see that the burden of knowledge channel explored in this paper asks more of other mechanisms if we wish to preserve the possibility of growth without an increase in research effort. The parametric independence of the burden of knowledge channel also leads us to specific empirical predictions that are independent of other stories. These predictions are defined in the next two sections.

3.7 Time Series Predictions

In addition to its predictions for the evolution of specialization (equation (25)), the model makes explicit predictions regarding the amount of education innovators seek and their propensity to form teams.

Consider education first. Since education is valuable to innovators and this value is complementary to growing income possibilities in the innovative sector – due to increasing market size if nothing else – innovator cohorts will seek more education over time. In equilibrium the optimal amount of education (equation (21)) is a fixed fraction of the innovator's lifetime income. As the economy grows, individual incomes grow at rate g. In consequence, the amount of education innovators seek also grows at rate g.

$$g_{E^*} = g \tag{26}$$

This is seen formally by taking logs in (21), differentiating with respect to time, and noting that v grows at rate g. The reason educational attainment grows at exactly the same rate as per-capita income comes from the Cobb-Douglas nature of the innovator's choice problem. The innovator pays an additive cost to acquire b, the breadth of expertise, which is an isoelastic input to the innovator's "production function", v. As is well known from the Cobb-Douglas case, the expenditure share on the input is a constant fraction of the income. Hence, as the income grows, the expenditure on the input grows at an equivalent rate.

Note that $\beta > 0$ is a necessary condition for growth in education expenditures in this model. If education (specifically, the breadth of expertise) were not valuable to innovators, then they would have no motive to seek education at all, let alone an increasing amount. Furthermore, the tendency to increase educational attainment over time implies that the breadth of expertise will increase in the absence of knowledge accumulation. This provides intuition for why $\varepsilon\delta > 1$, rather than the weaker condition $\varepsilon\delta > 0$, is required for the breadth of expertise to decline on the growth path.

Next consider the evolution of average team size. Recall that k, the breadth of expertise required to implement an idea, has a smooth distribution function $F(k)$. Recall also that teams are formed within cohorts if possible and that teams are formed with the minimum possible number of individuals (see Section 3.4). Since individuals allocate themselves evenly around the circle in any cohort, any necessary teammates are always available within one's own cohort. This implies that teams are formed from individuals with identical choices of b, $b^*(\tau)$. Since teams are formed from the minimum number of individuals, the implementation of any idea k requires $\lceil(k/b)$ team members; that is, k/b rounded up to the nearest integer.

The calculation of average team size is straightforward. A cohort with breadth of expertise b will produce a team of size 1 with probability $F(b)$, a team of size 2 with probability $F(2b) - F(b)$, and so on. The maximum team size in a cohort with breadth of expertise b is defined by $n = \lceil(1/b)$. With a little algebra, it is easy to show that expected team size is,[18]

$$\overline{team}(b) = n - \sum_{j=1}^{n-1} F(jb) \tag{27}$$

where j indexes a particular realization of team size. Differentiating (27) with respect to b shows that,

$$\frac{d\overline{team}(b)}{db} = -\sum_{j=1}^{n-1} jf(jb) \tag{28}$$

where $f(k) = dF/dk$ is the probability density function corresponding to F. Given that a density function is weakly positive at any point, we see that team size is weakly increasing as b falls. This result should seem intuitive: more specialized workers rely more on teamwork for the implementation of their ideas.[19]

[18] The expected team size is $\overline{team}(b) = 1F(b) + 2(F(2b) - F(b)) + ... + n(1 - F((n-1)b)$
Canceling terms in this expression produces the expression in the text.

[19] One might wonder about a more general case where the distribution F is parameterized by the breadth of expertise, b. In particular, we might imagine that more narrowly educated individuals will have a narrower range of inspiration (smaller average k). I explore this possibility formally in the Appendix and derive there a generalized condition for team size to increase as specialization increases. The intuition, which is shown clearly for a uniform distribution, is that team size will increase with specialization as long as the "reach" of innovators does not decline as rapidly as their "grasp". See the Appendix for details.

Finally, consider the time lag between two innovations in which the same innovator is involved. Given that individuals in a cohort of measure l each produce innovations with hazard rate λ, the cohort will produce in expectation $l\lambda dt$ innovations in an interval dt. The number of innovators needed to implement these innovations will be $(l\lambda dt)\,\overline{team}(b)$, while the supply of innovators is l. The probability that a single innovator will become involved in a project is therefore $(\lambda dt)\,\overline{team}(b)$, so that the individual's hazard rate is $h = \lambda\overline{team}(b)$. The average lag we witness is not $1/h$ however, but must consider the possibility that the innovator dies and so no additional innovation occurs. Given an innovation at time t, the expected lag conditional on witnessing an additional innovation before death is:[20]

$$\overline{lag} = \frac{\int_t^\infty h\widetilde{t}e^{-(h+\phi)(\widetilde{t}-t)}d\widetilde{t}}{\int_t^\infty he^{-(h+\phi)(\widetilde{t}-t)}d\widetilde{t}} = \frac{1}{\lambda\overline{team}(b)+\phi} \tag{29}$$

Given λ, the expected lag decreases as specialization and hence team size increase. As people become more specialized, they rely on each other more for the implementation of their ideas; there are more innovative opportunities for each person – less dead time waiting for a project – and the time lag between innovations drops. More generally, the evolution of the raw arrival rate of innovative ideas, λ, may reinforce or overturn the implications of increasing specialization. Since the evolution of λ along the growth path is ambiguous in the model, the model makes no simple prediction about the evolution in the lag. At the same time, should we find that increasing team size is not related to decreasing lags, the model suggests that the raw arrival rate of ideas must be declining; this interpretation will be helpful when we consider the empirical results.

In sum, the model predicts that (1) education is increasing over time; (2) specialization is increasing over time iff $\varepsilon\delta > 1$, and (3) team size is increasing over time iff $\varepsilon\delta > 1$. The model makes no simple prediction regarding the time lag between an innovator's innovations.

3.8 Cross-sectional Predictions

In this section I extend the model to consider variations across technological areas. The extension considers J unit circles of knowledge in place of a single circle. I assume that the elasticity parameters are the same across all areas of knowledge, while each circle has a

[20] Given that you have innovated at time t, the probability that you neither innovate again nor die by time $\widetilde{t}$ is $e^{-(h+\phi)(\widetilde{t}-t)}$, and the hazard rate of innovating at any time $\widetilde{t}$ is h. The numerator of (29) is the probability-weighted sum of possible time lags until the next successful innovation. The denominator is the probability of having another successful innovation (i.e., innovating again before death).

specific depth of knowledge D_j and a separate parameter A_j, which represents the relative productivity of knowledge in that area – whether the area is hot or cold. The structure of the model is as before, with two modifications. First, the difficulty of reaching the knowledge frontier will differ across technological areas. The educational cost for each area j is:

$$E_{ij}(\tau) = (b_i D_j(\tau))^{\varepsilon} \tag{30}$$

Second, an innovator's productivity will depend on the characteristics of the technological area. I redefine θ as

$$\theta_{ij}(t,\tau) = A_j(t) X(\tau)^{\chi} L_j(t, s_{ij})^{-\sigma} b_{ij}^{\beta} \tag{31}$$

This specification differs in two ways from that in equation (12). First, the congestion effects are now specific to the particular technological area, which is indicated by adding the subscript j to $L(t, s_i)$. Second, I add the new parameter, $A_j(t)$, to indicate sector specific research opportunities. Innovator's inspirations are drawn from a distribution $F_j[s_{ij}, s_{ij}+1]$, so that all ideas from an innovator operating in area j are implementable using expertise within that circle of knowledge.

The innovator's maximization problem is solved just as in Section 3.5, only we now consider the choice problem within a particular area of knowledge j. Congestion externalities imply that innovators evenly array themselves within any circle of knowledge, and the first order condition for b_{ij}^* becomes:

$$\frac{\beta}{b_{ij}^*} C X(\tau)^{\chi} b_i^{*\beta} \int_{\tau}^{\infty} A_j(t) L_{Rj}(t)^{-\sigma} L_Y(t) e^{-\phi(t-\tau)} dt = \frac{\varepsilon}{b_{ij}^*} \left(b_{ij}^* D_j(\tau)\right)^{\varepsilon} \tag{32}$$

Allowing $A_j(t)$ to grow at a sector specific rate, g_{Aj}, we find the following three results:

$$b_j^*(\tau) = \frac{1}{D_j(\tau)} \left(\frac{v_j(\tau)}{\phi - (1-\sigma) g_L - g_{Aj}} \frac{\beta}{\varepsilon} \right)^{1/\varepsilon} \tag{33}$$

$$E_j^*(\tau) = \frac{v_j(\tau)}{\phi - (1-\sigma) g_L - g_{Aj}} \frac{\beta}{\varepsilon} \tag{34}$$

$$U_j^*(\tau) = \frac{v_j(\tau)}{\phi - (1-\sigma) g_L - g_{Aj}} \left(1 - \frac{\beta}{\varepsilon}\right) \tag{35}$$

The central cross-sectional implications are seen directly. Income arbitrage across sectors implies that the $U_j^*(\tau) = U^*(\tau)\ \forall j$. This in turn implies that $E_j^*(\tau) = E^*(\tau)\ \forall j$. In other words, regardless of the depth of knowledge in a given sector or the innovation opportunities there, innovators will seek the same amount of education. The intuition for this result is that innovators will allocate themselves across sectors such that differences in the degree of congestion will offset the variation in technological opportunities or educational burden. Once income is equated across sectors, innovators acquire the same total education because their optimal amount of education is a constant fraction of their expected income. The model thus makes the perhaps surprising dual prediction that successive cohorts of innovators will choose an increasing amount of education, while a given cohort will choose an identical amount of education, regardless of difference in costs and opportunities across sectors.

Finally, while we expect no variation in the level of education across sectors, we do expect differences in specialization. Given income arbitrage, we can compare the specialization decisions across two sectors, j and j'. Using (33) we see directly that,

$$\frac{b_j^*(\tau)}{b_{j'}^*(\tau)} = \frac{D_{j'}(\tau)}{D_j(\tau)} \tag{36}$$

Specialization will be greater where the depth of knowledge is greater. In consequence, team size will also be greater where the depth of knowledge is greater.

In sum, the model predicts in cross-section that: (1) there will be no difference in the amount of education; (2) specialization will be greater where the depth of knowledge is greater; (3) team size will be greater where the depth of knowledge is greater.

4 Econometric Evidence

Sections 3.7 and 3.8 motivate a number of investigations. The goal of the empirical work is descriptive: to examine a range of first-order facts that, together, shed light on these predictions and the model's underlying parameters. Using an augmented patent data set, we will be able to examine four outcomes in particular:

1. Team size

2. Age at first innovation

3. Specialization, and

4. The time lag between innovations

The data is described in the following subsection. An investigation of basic time trends and cross-sectional results follow. The section closes by considering these new results together with the existing facts summarized in Section 2. Together they paint a multi-dimensional picture that is consistent with a rapidly increasing burden of knowledge.

4.1 Data

I make extensive use of a patent data set put together by Hall, Jaffe, and Trajtenberg (Hall et al. 2001). This data set contains every utility patent issued by the United States Patent and Trademark Office (USPTO) between 1963 and 1999. The available information for each patent includes: (i) the grant date and application year, and (ii) the technological category. The technological category is provided at various levels of abstraction: a 414 main patent class definition used by the USPTO as well as more organized 36-category and 6-category measures created by Hall et al. (The 36-category and 6-category measures are described in Table 4.5.) For patents granted after 1975, the data set includes additionally: (iii) every patent citation made by each patent, and (iv) the names and addresses of the inventors listed with each patent. There are 2.9 million patents in the entire data set, with 2.1 million patents in the 1975-1999 period. See Figure 4.1.

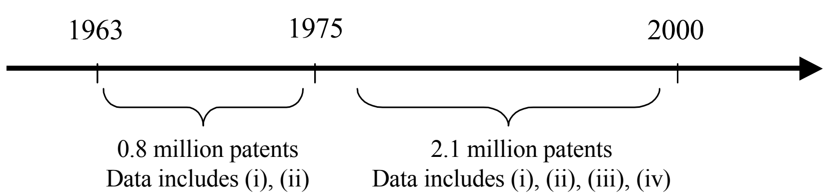

Figure 4.1: Summary of Available Data

Using the data available over the 1975-1999 time period, we can define two useful measures directly:

- Team Size. The number of inventors listed with each patent.
- Time Lag. The delay between consecutive patent applications from the same inventor.

For the latter measure, we identify inventors by their last name, first name, and middle initial and then build detailed patent histories for each individual.

We can also define two more approximate measures that will be useful for analysis:

- Tree Size. The size of the citations "tree" behind any patent. Any given patent will cite a number of other patents, which will in turn cite further patents, and so on. For the purposes of cross-sectional analysis, the number of nodes in a patent's backwards-looking patent tree serves as a proxy measure for the amount of underlying knowledge.

- Field Jump. The probability that an innovator switches technological areas between consecutive patent applications. This can serve as a proxy measure for the specialization of innovators. The more specialized you are, the less capable you are of switching fields.

A limitation of this last measure is that, since technological categories are assigned to patents and not to innovators, inferring an innovator's specific field of expertise is difficult when innovators work in teams. For inventors who work in teams, the relation between specialization and field jump is in fact ambiguous: as inventors become more specialized and work in larger teams, they may jump as regularly as they did before. For the specialization analysis we will therefore focus on solo inventors, for whom increased specialization is always associated with a decreased capability of switching fields.

Finally, we would like to investigate the age at first innovation. Unfortunately, inventors' dates of birth are not available in the data set, nor from the USPTO generally. However, using name and zip code information it was possible to attain birth date information for a large subset of inventors through a public website, www.AnyBirthday.com. AnyBirthday.com uses public records and contains birth date information for 135 million Americans. The website requires a name and zip code to produce a match. Using a java program to repeatedly query the website, it was found that, of the 224,152 inventors for whom the patent data included a zip code, AnyBirthday.com produced a unique match in 56,281 cases. The age data subset and associated selection issues are discussed in detail in the Data Appendix. The analysis there shows that the age subset is not a random sample of the overall innovator population. This caveat should be kept in mind when examining the age results, although it is mitigated by the fact that the differences between the groups become small when explained by other observables, controlling for these observables in the age regressions has little effect, and the results for team size, specialization and time lag persist when looking in the age subset. See the discussion in the Data Appendix.

4.2 Time series results

I consider the evolution over time of our four outcomes of interest. Figure 4.2 presents the basic data while Tables 4.1 through 4.4 examine the time trends in more detail.

Consider team size first. The upper left panel of Figure 4.2 shows that team size is increasing at a rapid rate, rising from an average of 1.70 in 1975 to 2.25 at the end of the period, for a 32% increase overall. Table 4.1 explores this trend further by performing regressions relating team size to application year, and we see that the time trend is robust to a number of controls. Controlling for compositional effects shows that any trends into certain technological categories or towards patents from abroad have little effect. Repeating the regressions separately for patents from domestic versus foreign sources shows that the domestic trend is steeper, though team size is rising substantially regardless of source. Repeating the time trend regression individually for each of the 36 different technological categories defined by Hall et al. shows that the upward trend in team size is positive and highly significant in every single technological category. Running the regressions separately by "assignee code" to control for the type of institution that owns the patent rights shows that the upward trend also prevails in each of the seven ownership categories identified in the data, indicating that the trend is robust across corporate, government, and other research settings, both in the U.S. and abroad.[21] In short, we find an upward trend in team size that is both general and remarkably steep.

Next consider the age at first innovation. Note that we define an innovator's "first" innovation as the first time they appear in the data set. Since we cannot witness individuals' patents before 1975, this definition is dubious for (i) older individuals, and (ii) observations of "first" innovations that occur close to 1975. To deal with these two problems, I will limit the analysis to those people who appear for the first time in the data set between the ages of 25 and 35 and after 1985. The upper right panel of Figure 4.2 plots the average age over time, where we see a strong upward trend. The basic time trend in Table 4.2 shows

[21] Table A.2 describes the ownership assignment categories.

an average increase in age at a rate of 0.66 years per decade. Controlling for compositional biases due to shifts in technological fields or team size has no effect on the estimates. The results are also similar when looking at different age windows.[22] Analysis of trends within technological categories shows that the upward trend in age is quite general. Smaller sample sizes tend to reduce significance when the data is finely cut, but an upward age trend is found in all 6 technology classes using Hall et al's 6-category measure, and in 29 of 36 categories when using their 36-category measure. The upward age trend also persists across all patent ownership classifications.

Now we turn to specialization. The specialization measure considers the probability that an innovator switches fields between consecutive innovations. Before looking at the raw data, it is necessary to consider a truncation problem that may bias us toward finding increased specialization over time. The limited window of our observations (1975-1999) means that the maximum possible time lag between consecutive patents by an innovator is largest in 1975 and smallest in 1999. This introduces a downward bias over time in the lag between innovations. It is intuitive, and it turns out in the data, that people are more likely to jump fields the longer they go between innovations.[23] Mechanically shorter lags as we move closer to 1999 can therefore produce an apparent increase in specialization. To combat this problem, I make use of a conservative and transparent strategy. I restrict the analysis to a subset of the data that contains only consecutive innovations which were made within the same window of time. In particular, we will look only at consecutive innovations when the second application comes within 3 years of the first. Furthermore, we will look only at innovations which were granted within 3 years of the application.[24] This strategy eliminates the bias problem at the cost of limiting our data analysis to the 1975-1993 period and making our results applicable only to the sub-sample of "faster" innovators.[25] The

[22] The table reports results for the 23 to 33 age window as well. In results not reported, I find that the trend is similar across subsets of these windows: ages 23-28, 25-30, 31-35, et cetera. Furthermore, there is no upward trend when looking at age windows beginning at age 35.

[23] An interpretation consistent with the spirit of the model is that people need time to reeducate themselves when they jump fields, hence a field jump is associated with a larger time lag.

[24] Looking only at patents where the second application came within 3 years limits our analysis to those cases where the first application was made before 1997. However, a second issue is that patents are granted with a delay – 2 years on average – and only patents that have been granted appear in the data. For a first patent applied for in 1996, it is therefore much more likely that we will witness a second patent applied for in 1997 than one applied for in 1999 – introducing further downward bias in the data. To deal completely with the truncation problem, we will therefore further limit ourselves to patents which were granted within 3 years of their application, which means that we will only look at the period 1975-1993.

[25] These restrictions maintain a significant percentage of the original sample. For example, of the 111,832

lower right panel of Figure 4.2 shows the trend from 1975-1993.

Table 4.3 considers the trend in specialization with and without this corrective strategy. The results there, together with the graphical presentation in Figure 4.2, indicate a smooth decrease in the probability of switching fields. The decline is again quite steep. Using the central estimate for the trend of -.003, we can interpret a 6% increase in specialization every ten years. Note that our main results, and Figure 4.2, use the 414-category measure for technology to determine whether a field switch has occurred. This is our most accurate measure of technological field (Hall et al.'s measures are aggregations of it), but the results are not influenced by the choice of field measure. Note in particular that the *percentage* trend is robust to the choice of the 6, 36, or 414 category measure for technology – the trend is approximately 6% per decade for all three. Including controls for U.S. patents, the application time lag, ownership status, and the technological class of the initial patent has little effect. Furthermore, looking for trends within each of Hall et al.'s 36 categories, we find that the probability of switching fields is declining in 34 of the 36; the decline is statistically significant in 20. In sum, we see a robust and strongly decreasing tendency for solo innovators to switch fields.

Finally, I consider the time lag between an innovator's innovations. The truncation bias in the time lag described above, which had little effect with specialization, is of course crucial here, so we employ the same corrective strategy and look only at the 1975-1993 period and the sub-sample of "faster" innovators. The lower left panel of Figure 4.2 presents the data graphically and Table 4.4 considers the trend with and without various controls. The regressions show a mild upward trend, but this should be viewed skeptically given the clearly cyclical behavior we see in the graph. Considering the coefficients on various controls, we see that bigger teams innovate faster and that part of the mild upward trend is accounted for by a composition effect – innovators switching into fields where the delay is longer. What is most interesting about the time lag data becomes apparent only when we look at trends within technological categories. (See Figure 4.3.) Here we find a richer story: Most fields (19 of 36) show a significant *decrease* in the average lag between innovations. A smaller number (11 of 36) show a significant increase.[26] Overall, I conclude that the

people who applied successfully for patents in 1975, 81,955 of them received a second patent prior to 2000. Of these 81,955 people for whom we can witness a time lag between applications, 79.8% made their next application within three years. Of those, 88.5% were granted both patents within three years of application.

[26] The fact that the overall trend is upward indicates that this group of 11 is pulling relatively strongly.

average time lag between an innovator's patent applications, unlike the other outcomes of interest, shows no decisive trend; rather, trends in time lags are cycling and differ strongly across technological areas.

4.3 Cross-section results

For a first look at the data in cross-section, Table 4.5 presents a simple comparison of means across the 6 and 36 technological categories of Hall et al (2001). The middle column in the table presents the mean age at first innovation, and the data shows a remarkable consistency across technological categories. In 30 of the 36 categories, an innovator's first innovation tends to come at age 29. The lowest mean age among the 36 categories is 28.8, and the highest is 31.1, though this last relies on only 12 observations and is an outlier with regard to the others. The table shows that regardless of whether the invention comes in "Nuclear & X-rays", "Furniture, House Fixtures", "Organic Compounds", or "Information Storage", the mean age at first innovation is nearly the same. According to the cross-sectional variation of the model, this is what we would expect. Given income arbitrage, innovators expand their breadth of expertise in shallow areas of knowledge and focus their breadth of knowledge in deep areas of knowledge so that their educational investment does not differ across fields.[27]

The next columns of the table consider the average team size. Here we see large differences across technological areas. The largest average team size, 2.90 for the "Drugs" subcategory, is over twice that of the smallest, 1.41 for the "Amusement Devices" subcategory.

Finally, the last columns of the table consider the probability that a solo innovator will switch sub-categories between innovations. Here, as with team size and unlike the age at first innovation, we see large differences across technological areas. This variation is again consistent with the predictions of the model. At the same time, this basic, cross-sectional

Upon closer examination we find that the heavyweights among these eleven are Organic Compounds (#14), Drugs (#31), and Biotechnology (#33) – all areas related to the pharmaceutical industry. This result is consistent with Henderson & Cockburn's (1996) finding that researchers in the pharmaceutical industry are having a greater difficulty in producing innovations over time.

[27] These results can also be considered in a regression format. Pooling cross-sections and using application year dummies to take care of trends, the results are extremely similar. One can also adjust the time at first innovation by subtracting category-specific estimates of the time lag to get a closer estimate of an individual's education. One can also look at different age windows. The result that ages are nearly identical across fields is highly robust.

variation in the probability of field jump is difficult to interpret: the probability of field jump will be tied to how broadly a technological category happens to be defined, which may vary to a large degree across categories.

I can go further by using a direct measure of the quantity of knowledge underlying a patent. In particular, I can analyze in cross-section what an increase in the knowledge measure implies for our outcomes of interest.

For a continuous measure of the quantity of knowledge I will use the logarithm of the number of nodes (i.e., patents) in the citation "tree" behind any patent.[28] As usual, there is a truncation issue that needs to be considered: the data set does not contain citation information for patents issued before 1975, so we tend to see the recent part of the tree. The measure of underlying knowledge is then noisier the closer we are to 1975, and I will therefore focus on cross-sections later in the time period. A second issue is that the average tree size and its variance grow extremely rapidly in the time window, which makes it difficult to compare data across cross-sections without a normalized measure. Two obvious normalizations are: (1) a dummy for whether the tree size is greater than the within-period median; (2) the difference from the within-period mean tree size, normalized by the within-period standard deviation. Results are reported using the latter definition, as it is informationally richer, though either method shows similar results.

Figure 4.4 presents, by application year, a set of kernel regressions relating the team size to the normalized variation in tree size. We see a very consistent pattern: a "J" shape. After a slight initial fall, team size rises at an increasing rate as the measure of knowledge depth increases. For innovations with larger citation trees, the rise in team size is particularly strong. At the right end of the figures, an increase of one standard deviation in the tree size is associated with an average increase in team size of one person.[29]

[28] The distribution of the raw node count within cross-section is highly skewed – the mean is far above the median, so that upper tail outliers can dominate the analysis. I therefore use the natural log of the node count, which serves to contain the upper tail. A (loose) theoretical justification is knowledge depreciation: distant layers of the tree are less relevant to a patent than nearer layers, so there is a natural diminishing impact as nodes grow more distant. The diminishing impact of the large, distant layers, which dominate the node counts, is captured loosely by taking logs. Noting that the basic results are similar when we use the median-based measure of knowledge depth (a dummy for whether the raw node count is above or below the median, which is independent of any monotonic transform of the node count) we can be reasonably comfortable with the log measure.

[29] While the rising relationship is consistent with the predictions of the model, the slight initial fall is not. Re-examining the relationship between average team size and average tree size by technological category shows a surprising fact, which can be seen in Figure 4.5. There are a few technological fields that have high team size but small citation trees. These outliers are: Organic Compounds (#14), Drugs (#31),

Table 4.6 reexamines the relationship between team size and tree size in pooled cross-sections, with and without various controls. I add a quadratic term for the variation in team size to help capture the curvature seen in the figures.[30] The table shows that the cross-sectional relationship holds for domestic and foreign-source patents and when controlling for technological category, so that the variation appears both within fields and across them. Technological controls are perhaps best left out, however, since the variations in mean tree size across technological category may be equally of interest. Finally, we might be concerned that bigger teams simply have a greater propensity to cite, which results in larger trees. This concern proves unwarranted. Controlling for the variation in the direct citations made by each patent, we find that relationship actually strengthens. In fact, we see that bigger teams tend to cite *less*. This result gives us greater faith in the causative arrow implied by the regressions.

Next we turn to the age at first innovation. Table 4.7 examines, in pooled cross-sections, the relationship between age and knowledge for those individuals for whom we can be confident that they are innovating for the first time (see discussion above). The general conclusion from the table is that we must work hard to find a relationship, and at its largest it is very small. It is not robust to the specific age window, is reduced when controlling for the technological category, and disappears when controlling for the number of direct citations made. Taking a coefficient of 0.1 as the maximum estimate from the table, we find that an increase of one standard deviation in the knowledge measure leads to a 0.1 year increase in age. This coefficient may be attenuated given that our proxy measure of knowledge is, at best, noisy, but I conclude that there is at most only a weak relationship between the amount of knowledge underlying a patent and the age at first innovation.

Finally, Table 4.8 considers the relationship between the probability of field jump and the knowledge measure. The table shows a robust negative relationship: solo innovators

and Biotechnology (#33). Interestingly, these are exactly the same fields which were dragging strongly upwards on the trend in the time lag between applications, examined above. A plausible explanation for the unusual behavior of these three categories is the move from "random" to "rational" research and the consequent increasing need for specialists in the pharmaceutical industry, which is described by Henderson & Cockburn (1996) and others. A tendency toward random discovery in the past will provide little prior art for innovations and result in small citation trees. The growing need for specialists will result in large teams.

[30] The increase in slope is intriguing, but difficult to interpret, since the tree size is a proxy measure for the amount of knowledge underlying a patent. Interpretations are further complicated by the fact that the curvature will change depending on the monotonic transform we use for the tree size (in this case, the logarithm).

are less likely to jump fields when their initial patent has a larger node count. If we identify a larger node count with a deeper area of knowledge, then this negative correlation is again consistent with the predictions of the model. However, I place less emphasis on this result. The fact that the node count captures the recent part of the tree means that the measure is likely correlated not just with the total underlying knowledge but also with the recent ease of innovation. This effect could also explain the negative correlation. Innovators will be less likely to leave a fruitful area, which will be registered as a decreased probability of jumping fields.

4.4 Interpretations

I have assembled a collection of new facts, motivated by the model. This section considers these facts as a whole to see whether they are consistent with the model and what they say about other models of growth more generally.

The model predicts that successive cohorts of innovators will seek more education, because education is valuable to innovators and its value increases as the economy grows. The model also predicts that, due to income arbitrage, innovators will seek the same amount of education across widely different areas of knowledge, regardless of variations in the depth of knowledge or innovative opportunities. These dual predictions find strong support in the data.

Second, the model indicates that if knowledge is accumulating at a sufficient rate and education is sufficiently costly (so that $\varepsilon\delta > 1$), then innovators will seek a greater degree of specialization over time. Increasing specialization will result in greater teamwork as innovators become more interdependent in the implementation of their ideas. In cross-section, specialization and team size are predicted to vary across fields and, in particular, to be greater where the depth of knowledge is greater. These time-series and cross-sectional predictions all find consistent empirical support. With an increasing burden of knowledge, the model indicates pessimistic predictions for growth, as were discussed in Section 3.6.

How does a story of increasing knowledge burden do with the data aggregates presented in Section 2? First of all, if the knowledge burden is rising, we might wonder why there are more and more people engaging in R&D (see Figure 2.1). The rise in research effort is natural, however, given the increase in market size – the value of patents is increasing on the extensive margin. This market size effect is present in this model as in other idea-based growth models. More interesting is the drop in patent production per active researcher. Figure 2.2 shows the recent trend, but the fact of declining patent output per researcher may date back as far as 1900 and even before (Machlup 1962). Certainly, not all researchers are engaging in patentable activities, and it is possible that much of the trend is explained by a relatively rapid growth of research in basic science.[31] Still, it is quite interesting to note that the recent drop in patents per U.S. R&D worker, a drop of about 50% since 1975, is roughly consistent in magnitude with the rise in U.S. team size over that period. With the time lag between innovations showing little if any deterministic trend, we have a simple explanation for where these extra innovators have recently been going – into bigger teams.

What of other stories? As emphasized in Section 3.6, the knowledge burden channel is not dispositive of other mechanisms, which operate independently and may also be important to growth. At the same time, it is worth considering briefly whether popular stories in the growth literature can serve as alternate explanations for the facts collected in this paper.

First, models that explain away scale effects through an expanding product space cannot obviously explain many of these facts. Expanding team size and rising ages at first innovation seem outside their predictive thrust. Moreover, as noted in Section 2, they do not on their own explain the declining number of patents per researcher.

Models that avoid scale effects through the evolution of the quality of an innovator's ideas (i.e. "fishing out" type stories) can do well with the data aggregates, but they do not provide obvious first-order explanations for the team size, age, or specialization data. At the same time, there may be some indirect evidence that successful ideas are in fact harder to come by. To see this, consider that, given a fixed arrival rate of ideas λ, an increase in specialization in our model predicts a decrease in the time lag between innovators' innovations: innovators become more interdependent so that they share in larger numbers of projects.[32] In the data, innovators in most technological classes do show a decrease in the lag over time, but a large minority of technological categories show an upward trend in time lag, and the overall behavior is cyclical. Why don't we see a strong drop in the time lag? One explanation

[31] Such an explanation could be inferred from the observations of Mokyr (1990), for example, who sees an increasing role for basic science as a foundation for technological advance.

[32] Furthermore, although the model abstracts from implementation time, we could also imagine that a given project would be implemented faster when more people are brought to bear. This effect would also tend to reduce the time lag between an innovator's inventions.

is that the underlying arrival rate of ideas, λ, may be declining. From equation (10), we see three possible explanations for such a decline. First, this result is consistent with the negative impact of narrowing expertise on the frequency of an innovators ideas ($\beta_\lambda > 0$). In this sense, we need look no further than the knowledge burden mechanism to explain this result. However, it is also consistent with a "fishing out" problem ($\chi_\lambda < 0$). Finally, it is consistent with an increase in competition ($\sigma > 0$).

Competition effects may present a more complete alternative explanation for the range of data. If we think of the innovation process as a series of increasingly competitive patent races, we can explain the time lag results and go further as well. As market size grows, patents become more valuable. Competition within patent lines will increase and we will consequently see more "losers". This can explain an increasing time lag and a drop in patent counts per active innovator. Team sizes, insofar as teams reduce innovation time, may expand as firms try to out-race each other. With an increase in team size, we might see an increase in specialization to exploit within-team efficiency possibilities. The fact that we see increased specialization among *solo* inventors is less easy to explain however. It is also difficult to produce the age results through competition. The fact that the average age at first innovation is increasing over time but showing no variation across fields poses a particular challenge. As a general matter, increasing specialization despite increasing educational attainment is difficult to reconcile without appealing to an increasing educational burden. Another problem with a patent race story, if it is to stand on its own, is that it must make extreme assumptions to reproduce the data aggregates in Section 2. Given historically flat TFP growth and historically flat patent counts, we must imagine that all the extra research effort is useless: no matter how many people enter R&D, the number of patent races is fixed and no race is resolved any faster.

5 Conclusion

If technological progress leads to an accumulation of knowledge, then the educational burden on successive generations of innovators will increase. Innovators may compensate by narrowing their expertise, which serves to reduce their individual capabilities, with negative implications for growth. This paper explores this possibility in a model that generates a number of empirical tests. The model predicts that the educational attainment of innovators will not vary across technological fields but will rise over time. Data analysis shows that the age at first innovation, which is a proxy measure for education, is in fact remarkably consistent across technological areas but is increasing over time at a rate of 0.6 years per decade. The model further predicts that specialization and average team size will vary across fields and, if the knowledge burden mechanism is sufficiently strong, specialization and team size will increase as technology advances. Data analysis shows that specialization and team size do vary considerably across technological areas and both show a sharp increase over time; specialization is increasing by 6% per decade, U.S. team size by 17% per decade. Furthermore, in cross-section, specialization and team size are positively correlated, as predicted, with a direct measure of the amount of knowledge underlying each patent. The knowledge burden mechanism thus provides a consistent explanation for the range of new evidence. It can also explain the facts standing at the center of the "scale effects" debate in the growth literature: flat patent counts and flat TFP growth despite rising R&D effort. The implication of these facts, understood through the lens of a rising burden of knowledge, is that growth is relying on an ever increasing and possibly unsustainable rise in innovative effort.

This paper aims to provide a deeper understanding of key issues in innovation and growth and particularly to introduce a specific and isolated channel – an increasing educational burden – through which innovation can become harder with time. Further work should extend the empirical explorations over longer periods of time and, if possible, produce more closely identified tests for this mechanism and others. Detailed modeling of the microfoundations may suggest further approaches.

6 Appendix

Proof that $L(t,s) = L_R(t)\ \forall s$

This proof proceeds in two steps. First I rule out any equilibrium in which there is zero mass at a proper subset of points on the circle. Then I show, given the first result, that innovators will array themselves evenly around the circle.

(1) $L_R(\tau)$ is the mass of individuals at time τ who are engaged in R&D. Define G_τ as the set of points on the circle where there is a positive mass of innovators:

$$G_\tau = \{s \in [0,1] \mid L(\tau,s) > 0\}$$

The set of points where there is zero mass is defined as the complement of G_τ, $H_\tau = G_\tau^c$. If G_τ is empty, then $L(\tau,s) = 0\ \forall s$, which satisfies the proof trivially. Consider the more interesting and relevant case where G_τ is non-empty. I first prove that H_τ must then be empty.

By contradiction, assume the set H_τ is non-empty, $H_\tau \neq \emptyset$. Then there exists at least one point s' on the boundary between G_τ and H_τ where for any $\epsilon > 0$ there exists a point s'' such that either (i) $s' \in G_\tau$, $s'' \in H_\tau$, $s' > s'' > s' - \epsilon$, or (ii) $s' \in H_\tau$, $s'' \in G_\tau$, $s' < s'' < s' + \epsilon$. Consider the former case.[33] For ease of exposition, I will abuse notation slightly and let $L(t,s)$ represent both the set and the mass of individuals at point s at time t. Then there must exist some massless innovator $i \in L(\tau, s')$ with breadth of expertise b_i^* who chose position s' at some time $t' \leq \tau$. Without loss of generality, choose i such that $b_i^* \leq b_j$ for some $j \in L(t', s')$, $j \neq i$. Note further that $b_i^* > 0$ in any equilibrium, by the arbitrage condition (15), since $U_i^{R\&D}(b_i^* = 0) = 0$ but $U^{wage} > 0$; hence there must exist an arbitrarily small ϵ such that $0 < \epsilon < b_i^*$.

The boundary of this individual's knowledge is $s' + b_i^*$. If the innovator has an idea $k > b_i^*$, then the innovator will need teammates for implementation. Define $p_i(t)$ such that $\forall k > p_i(t)$ the necessary teammates do not exist and $\forall k \leq p_i(t)$ the necessary teammates do exist. The probability that an idea k is feasible is then $F(p_i(t))$. The innovator's expected income at the time of their birth t' is a generalized version of (6) that allows for the possibility that an idea is infeasible:

$$CX(t')^\chi b_i^{*\beta} \int_{t'}^{\infty} F(p_i(t)) L(t,s')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt - \left(b_i^* D(t')\right)^\varepsilon \tag{37}$$

If this individual were to shift to a location $s'' \in H_\tau \subset H_{t'}$, then the access to potential teammates remains unchanged. (The individual can always hire someone in $L(t', s')$ as a teammate, and everyone else at that point has weakly greater expertise.) Therefore $\widehat{p_i}(t) = p_i(t) + \epsilon$, and the probability that an idea k is feasible is weakly increasing since for any distribution function $F(p_i(t) + \epsilon) \geq F(p_i(t))$.

Therefore, from (37) and the equilibrium condition (14), the choice s' can only be an equilibrium for person i if

$$\int_{t'}^{\infty} L(t,s')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt \geq \int_{t'}^{\infty} L(t,s'')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt \tag{38}$$

[33] The proof for case (ii) follows on similar lines; I omit it for brevity.

Given the continuity of L with time, $L(t, s') > L(t, s'')$ for all t in some interval $[t', t'']$. Therefore, the expected income to innovator i in the interval $[t', t'']$ must be strictly less with the choice s' then with the choice s''. Therefore, innovator i must believe that

$$\int_{t''}^{\infty} L(t, s')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt > \int_{t''}^{\infty} L(t, s'')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt \quad (39)$$

Multiplying both sides of this expression by the constant $e^{-\phi(t'-t'')}$, we see that in the subgame for those born at time t'' no person would choose s''. Hence $L(t'', s'')$ is not increasing, $L(t'', s') > L(t'', s'')$, and there is no finite t'' at which (39) holds. Hence (38) cannot hold. Hence, by contradiction, no such point s'' can exist and therefore $H_\tau = \emptyset \; \forall \tau$.

(2) Given that $H_\tau = \emptyset \; \forall \tau$, innovators' collective expertise covers all areas of knowledge, so that all ideas are feasible to implement. The proof that innovators array themselves evenly around the circle then follows as above. By contradiction, assume an innovator born at time t' chooses s' over some s'' where $L(t', s') > L(t', s'')$. The innovator must believe (38), and by extension (39). But there is no t'' at which (39) holds. Hence no such point s'' can exist. Hence no innovator can choose any s' such that $L(t', s') > L(t', s)$ for any s. QED

Derivation of the steady-state growth rate

From equation (23), the steady-state growth rate in the economy is defined by,

$$g = g_{\bar{\theta}} + g_L \quad (40)$$

To define $g_{\bar{\theta}}$, note first that the average productivity of innovators is the sum of the productivity of each cohort weighted by the fraction of that cohort in the population.[34]

$$\overline{\theta}(t) = \int_{-\infty}^{t} \theta(\tau, t)(g_L + \phi) e^{(g_L+\phi)(\tau - t)} d\tau \quad (41)$$

The growth rate of $\theta(\tau, t)$ with respect to τ is just $\chi g + \beta g_{b^*}$, which is seen by taking logs of the definition of θ (equation (12)), using the equilibrium result $b_i = b^*(\tau)$, and differentiating with respect to τ. We can therefore integrate (41) to find that $\overline{\theta}(t) = \theta(t)\,(g_L + \phi) \,/\, (\chi g + \beta g_{b^*} + g_L + \phi)$. The steady-state growth rate in $\overline{\theta}(t)$ is therefore equivalent to the steady-state growth rate in $\theta(t)$, so that $g = g_{\bar{\theta}} + g_L = g_\theta + g_L$.

$\theta(t)$ is the productivity of the latest cohort of innovators at the time of their birth. The growth rate of $\theta(t)$ with respect to time is just $\chi g + \beta g_{b^*} - \sigma g_L$, which is seen by taking logs in the definition of θ, letting $\tau = t$, and differentiating with respect to t. Therefore,

$$g_{\bar{\theta}} = \chi g + \beta g_{b^*} - \sigma g_L \quad (42)$$

Taking logs in the equilibrium result (20), letting $\tau = t$, and differentiating with respect to t, the growth rate of b^* is $g_{b^*} = (1/\varepsilon) g_v - g_D$. Noting from (13) that $v(t) = \theta(t) C L_Y(t)$,

[34] The size of a cohort at its birth is $(g_L + \phi)L(\tau)$, so the surviving size of that cohort at some time $t > \tau$ is $(g_L + \phi)L(\tau)e^{-\phi(t-\tau)}$, and the fraction of the population $L(t)$ who belong to that cohort is $(g_L + \phi)e^{-g_L(t-\tau)}e^{-\phi(t-\tau)}$.

the growth rate in v is the same as the growth rate in the economy: $g_v = g_\theta + g_L = g$. From (8), the growth rate in the depth of knowledge is $g_D = \delta g$. Therefore,

$$g_{b^*} = (\frac{1}{\varepsilon} - \delta)g$$

Inserting this into (42), the result into (40), and rearranging produces the expression for steady-state growth in equation (24).

A generalized condition for team size to increase with specialization

I explore here the evolution of team size when the distribution of k changes with an individual's breadth of expertise, b. Define the generalized distribution function by $F(k;b)$ and the corresponding density function as $f(k;b) = dF(k;b)/dk$. The average team size for a cohort with breadth of expertise b is derived just as in (27),

$$\overline{team}(b) = n - \sum_{j=1}^{n-1} F(jb;b) \tag{44}$$

Noting that $F(jb;b) = \int_0^{jb} f(k;b)dk$, we can use Leibniz's rule to differentiate (44) with respect to b and thereby define a necessary and sufficient condition for team size to increase with specialization:

$$\sum_{j=1}^{n-1} \left(\int_0^{jb} \frac{df(k;b)}{db} dk + jf(jb;b) \right) > 0 \Leftrightarrow \frac{d\overline{team}(b)}{db} < 0 \tag{45}$$

The second term on the left hand side is recognized from equation (28) and acts to make team size increase with specialization. The effect of the first term is ambiguous, however, so that the effect of specialization on team size cannot be signed without considering distribution-specific properties.

We can gain some intuition for this condition by considering the simple case where k is drawn from a uniform distribution. Specifically, let $k \sim U[0, b^\alpha]$, so that $f(k;b) = b^{-\alpha}$. Using (45), it is then straightforward to show that $\alpha < 1 \Leftrightarrow d\overline{team}(b)/db < 0$. Noting that the mean of k is $E(k) = \frac{1}{2}b^\alpha$, it is also straightforward to show that α is the elasticity of $E(k)$ with respect to b. Therefore, we see that team size will be increasing with specialization so long as the elasticity of $E(k)$ with respect to expertise is less than 1. In other words, team size will be increasing as long as innovators' average "reach", given by $E(k)$, does not decline faster than their average "grasp", given by b.

7 Data Appendix

The reader is referred to Hall et al (2001) for a detailed discussion of their patent data set. This appendix focuses on the age information collected to augment the Hall et al data.

Age data was collected using the website www.AnyBirthday.com, which requires a name and zip code to produce a match. As is seen in Table A.1, 30% of U.S. inventors listed a zip code on at least one of their patent applications, and of these inventors AnyBirthday.com produced a birth date in 25% of the cases. While the number of observations produced by AnyBirthday.com is large, it represents only 7.5% of U.S. inventors. This Appendix explores the causes and implications of this selection. The first question is why zip code information is available for only certain inventors. The second question is why AnyBirthday.com produces a match only one-quarter of the time. The third question is whether this selection appears to matter.

Table A.2 compares how patent rights are assigned across samples. The table shows clearly that zip code information is virtually always supplied when the inventor has yet to assign the rights; conversely, zip code information is never provided when the rights are already assigned. Patent rights are usually assigned to private corporations (80% of the time) and remain unassigned in the majority of the other cases (17% of the time). An unassigned patent indicates only that the inventor(s) have not yet assigned the patent at the time it is granted. Presumably, innovators who provide zip codes are operating outside of binding contracts with corporations, universities, or other agencies that would automatically acquire any patent rights. The zip-code subset is therefore not a random sample, but is capturing a distinct subset of innovators who, at least at one point, were operating independently. Despite this distinction, this subset may not be substantially different from other innovators: the last column of Table A.2 indicates that, when looking at the other patents produced by these innovators, they have a similar propensity to assign them to corporations as the U.S. population average.

The nature of the selection introduced by AnyBirthday.com is more difficult to identify. The website reports a database of 135 million individuals and reports to have built its database using "public records". Access to public records is a contentious legal issue.[35] Public disclosure of personal information is proscribed at the federal level by the Freedom of Information Act and Privacy Act of 1974. At the state and local level however, rules vary. Birth date and address information are both available through motor vehicle departments and their electronic databases are likely to be the main source of AnyBirthday.com's records.[36] The availability of birth date information is therefore very likely to be related to local institutional rules regarding motor vehicle departments. Geography thus will influence the presence of innovators in the age sample, and a further issue in selection may involve the geographic mobility of the innovator, among other factors. The influence of this selection, together with the implications of assignment status, can be assessed by comparing observable means in the population across subsamples.

Table A.3 considers average team size, which is a source of further differences. Patents with provided zip codes have smaller team sizes than the U.S. average; team sizes in the subset of these patents for which the age of one innovator is known are slightly larger, but

[35] Repeated requests to AnyBirthday.com to define their sources more explicitly have yet to produce a response.

[36] A federal law, the Driver's Privacy Protection Act of 1994, was introduced to give individuals increased privacy. The law requires motor vehicle departments to receive explicit prior consent from an individual before disclosing their personal information. However, the law makes an exception for cases where motor vehicles departments provide information to survey and marketing organizations. In that case, individual's consent is assumed unless the individual has opted-out on their own initiative. See Gellman (1995) for an in-depth discussion of the laws and legal history surrounding public records.

still smaller than the U.S. average. Controlling for other patent observables, in particular the assignment status, reduces the mean differences and brings the age sample quite closely in line with the U.S. mean. (See the last two columns of the table.) Having examined a number of other observables in the data, such as citations received and average tree size, I find that relatively small differences tend to exist in the raw data, and that these can be either entirely or largely explained by controlling for assignment status and team size. Most importantly, the age results in the text are all robust to the inclusion of assignment status, team size, and any other available controls.

Finally, looking at team size, specialization, and time lag trends in the age subsample, the results are similar in sign and significance as those presented in Section 4. The rate of increase in specialization is larger, and the rate of increase in team size is smaller. The time lag shows no trend. Reexamining trends in the entire data set by assignment status, I find that the team size trend is weaker among the unassigned category, which likely explains the weaker trend in the age subset. Similarly, I find that the specialization trend is stronger among the unassigned category, which likely explains the stronger trend in the age subset.

I conclude therefore that while the age subset is not a random sample of the U.S. innovator population, the differences tend to be explainable with other observables and, on the basis of including such observables in the analysis, the age results appear robust.

References

[1] Aghion, Philippe and Howitt, Peter. "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, 60, 323-351.

[2] ——. *Endogenous Growth Theory*, Cambridge, MA: MIT Press, 1998.

[3] Baily, Martin N. and Chakrabarti, Alok K. "Innovation and Productivity in U.S. Industry," *Brookings Papers on Economic Activity*, 1985, 609-632.

[4] Evenson, Robert E. "International Invention: Implications for Technology Market Analysis," in Zvi Griliches, ed., *R&D, Patents, and Productivity*, Chicago, IL: University of Chicago Press, 1984, 89-123.

[5] Gellman, Robert. "Public Records: Access, Privacy, and Public Policy," Center for Democracy and Technology Discussion Paper, April 1995, http://www.cdt.org/privacy/pubrecs/pubrec.html.

[6] Grossman, Gene M. and Helpman, Elhanan. *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press, 1991.

[7] Hall, Bronwyn H., Jaffe, Adam B., and Trajtenberg, Manuel. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," NBER Working Paper No. 8498, October 2001.

[8] Henderson, Rebecca and Cockburn, Iain. "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery," *Rand Journal of Economics*, Spring 1996, 27, 32-59.

[9] Jones, Charles I. "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics*, May 1995, 110, 495-525.

[10] ——. "R&D-Based Models of Economic Growth," *Journal of Political Economy*, August 1995, 103, 759-784.

[11] ——. "Growth: With or Without Scale Effects?" *American Economic Review Papers and Proceedings*, May 1999, 89, 139-144.

[12] Kremer, Michael. "Population Growth and Technological Change: One Million B.C. to 1990," *Quarterly Journal of Economics*, August 1993, 108, 681-716.

[13] Kortum, Samuel S. "Equilibrium R&D and the Decline in the Patent-R&D Ratio: U.S. Evidence," *American Economic Review Papers and Proceedings*, May 1993, 83, 450-457.

[14] ——. "Research, Patenting, and Technological Change," *Econometrica*, November 1997, 65, 1389-1419.

[15] Kuhn, Thomas. *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press, 1962.

[16] Machlup, Fritz. *The Production and Distribution of Knowledge in the United States*, Princeton, NJ: Princeton University Press, 1962, 170-176.

[17] Malone, Michael S. *The Microprocessor: A Biography*, New York, NY: Springer-Verlag New York, 1995.

[18] Mansfield, Edwin. "Patents and Innovation: An Empirical Study," *Management Science*, February 1986, 32, 173-181.

[19] Mokyr, Joel. *The Lever of Riches*, New York, NY: Oxford University Press, 1990.

[20] Romer, Paul M. "Endogenous Technological Change," *Journal of Political Economy*, October 1990, 98, S71-S102.

[21] Segerstrom, Paul. "Endogenous Growth Without Scale Effects," *American Economic Review*, December 1998, 88, 1290-1310.

[22] Weitzman, Martin L. "Recombinant Growth," *Quarterly Journal of Economics*, May 1998, 113, 331-360.

[23] Young, Alwyn. "Growth Without Scale Effects," *Journal of Political Economy*, February 1998, 106, 41-63.

Table 4.1: Trends in Inventors per Patent

		Dependent Variable: Inventors per Patent						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Application Year		.0293 (.0001)	.0261 (.0001)	.0262 (.0001)	.0251 (.0001)	.0244 (.0002)	.0306 (.0002)	.0180 (.0003)
Foreign Patent		--	.444 (.002)	.416 (.002)	.141 (.004)	.146 (.004)	US Only	Foreign Only
Technological Field Controls	Broad	--	Yes	--	--	--	--	--
	Narrow	--	--	Yes	Yes	Yes	Yes	Yes
Assignee Code		--	--	--	Yes	Yes	Yes	Yes
Number of Observations		2,016,377	2,016,377	2,016,377	2,016,377	1,506,956	1,123,310	893,067
Period		1975-1999	1975-1999	1975-1999	1975-1999	1975-1996	1975-1999	1975-1999
Mean of Dependent Variable		2.03	2.03	2.03	2.03	1.97	1.82	2.29
Per-decade Trend as % of Period Mean		14.4%	12.9%	12.9%	12.4%	12.4%	16.8%	7.9%
R^2		.02	.08	.10	.12	.13	.12	.10

NOTES

(i) Regressions are OLS with standard errors in parentheses. Specifications (1) through (4) consider the entire universe of patents applied for between 1975 and 1999. Specification (5) considers only patents that were granted within three years after application (see discussion in text). Specifications (6) and (7) present separate trends for domestic and foreign source patents.

(ii) Foreign Patent is a dummy variable to indicate whether the first inventor listed with the patent has an address outside the U.S..

(iii) "Broad" technological controls include dummies for each of the 6 categories in Hall et al.'s most aggregated technological classification. "Narrow" technological controls include dummies for each category of their 36-category classification.

(iv) Upward trends persist when run separately for each technological field. Using the broad classification (six categories), the trends range from a low of .018 for "Other" to a high of .037 for "Chemical". Using the narrower classification scheme (thirty-six categories), the trends range from a low of .007 for "Apparel &Textile" to .051 for "Organic Compounds". The smallest t-statistic for any of these trends is 7.76.

(v) Assignee code controls are seven dummy variables that define who holds the rights to the patent. Most patent rights are held by US or foreign corporations (80%); while a minority remain unassigned (17%) at the time the patent is issued. Table A.2 describes the assignee codes in further detail. Running the time trends separately for the individual assignee codes shows that the team size trends range from a low of .005 for the unassigned category to a high of .039 for US non-government institutions. The lowest t-statistic for any of these trends is 5.38.

Table 4.2: Trends in Age at First Innovation

		Dependent Variable: Age at application						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Application Year		.0657 (.0095)	.0666 (.0095)	.0671 (.0095)	.0671 (.0099)	.0687 (.0097)	.0530 (.0107)	.0584 (.0109)
Technological Field Controls	Broad	--	Yes	--	--	--	--	--
	Narrow	--	--	Yes	Yes	Yes	--	Yes
Assignee Code		--	--	--	Yes	Yes	--	Yes
Team Size		--	--	--	--	-.0630 (.0273)	--	-.0348 (.0306)
Number of observations		6,541	6,541	6,541	6,541	6,541	5,102	5,102
Period		1985-1999	1985-1999	1985-1999	1985-1999	1985-1999	1985-1999	1985-1999
Age Range		25-35	25-35	25-35	25-35	25-35	23-33	23-33
Mean of Dependent Variable		31.0	31.0	31.0	31.0	31.0	29.3	29.3
Per-decade Trend as % of Period Mean		2.1%	2.1%	2.2%	2.2%	2.2%	1.8%	2.0%
R^2		.007	.010	.020	.020	.021	.005	.018

NOTES

(i) Regressions are OLS, with standard errors in parentheses. All regressions look only at those innovators for whom we have age data and who appear for the first time in the data set in or after 1985. Specifications (1) through (5) consider those innovators who appear for the first time between ages 25 and 35. Specifications (6) and (7) consider those innovators who appear for the first time between ages 23 and 33.

(ii) "Broad" technological controls include dummies for each of the 6 categories in Hall et al.'s most aggregated technological classification. "Narrow" technological controls include dummies for each classification in their 36-category measure. The upward age trend persists when run separately in each of Hall et al's broad technology classes. These trends are significant in 5 of the 6 categories, with similar trend coefficients as when the data are pooled. Upward trends are also found in 29 of 36 categories when using Hall et al.'s narrow technology classification. Here 12 categories show significant upward trends. Sample sizes drop considerably when the data is divided into these 36 categories. The one case of a significant downward trend (category #23, Computer Peripherals) has 42 observations.

(iii) Assignee code controls are seven dummy variables that define who holds the rights to the patent. Table A.2 describes the assignee codes in further detail. The upward age trends persist when run separately for each assignee code and are similar in magnitude to the trends in the table above.

Table 4.3: Trends in Probability of Field Jump

	Dependent Variable: Probability of Switching Technological Field							
	(1) 414	(2) 414	(3) 36	(4) 36	(5) 6	(6) 6	(7) 414	(8) 414
Application Year	-3.4e-3 (.19e-3)	-3.2e-3 (.19e-3)	-2.5e-3 (.19e-3)	-2.8e-3 (.19e-3)	-1.9e-3 (.17e-3)	-2.3e-3 (.17e-3)	-5.1e-3 (.12e-3)	-3.0e-3 (.11e-3)
Foreign Patent	--	.0076 (.0039)	--	-.0041 (.0038)	--	.0002 (.0035)	--	-.0005 (.0029)
Time Between Applications	--	.0225 (.0012)	--	.0206 (.0012)	--	.0154 (.0011)	--	.0228 (.0004)
Technological Field Controls (first patent)	--	Yes	--	Yes	--	Yes	--	Yes
Assignee Code (first patent)	--	Yes	--	Yes	--	Yes	--	Yes
Number of observations	215,855	215,855	215,855	215,855	215,855	215,855	359,405	359,405
Period	1975-1993	1975-1993	1975-1993	1975-1993	1975-1993	1975-1993	1975-1999	1975-1999
Mean of Dependent Variable	.535	.535	.423	.423	.294	.294	.556	.556
Per-decade Trend as % of Period Mean	-6.4%	-6.0%	-5.9%	-6.4%	-6.5%	-7.8%	-9.4%	-5.6%
(Pseudo) R^2	.0011	.018	.0006	.019	.0005	.017	.004	.026

NOTES
(i) Results are for probit estimation, with coefficients reported at mean values and z-statistics in parentheses. The coefficient for the Foreign dummy is reported over the 0-1 range.
(ii) The dependent variable is 0 if an inventor does not switch fields between two consecutive innovations. The dependent variable is 1 if the inventor does switch fields. Column headings define the field classification used to determine the dependent variable: "414" indicates the 414-category technological class definition of the USPTO; "36" and "6" refer to the aggregated measures defined by Hall et al (2001).
(iii) Specifications (1) through (6) consider "fast" innovators -- only those consecutive patents with no more than 3 years between applications and with no more than 3 years delay between application and grant. (See discussion in text.) Specifications (7) and (8) consider all consecutive patents.
(iv) Technological field controls are dummies for the 36 categories defined by Hall et al (2001). The reported regressions use the technological field of the initial patent. Using the field of the second patent has no effect on the results. Running the regressions separately by technology category shows that the trends persist in 6 of 6 categories using Hall et al.'s broad technology classification and 34 of 36 categories using Hall et al's narrow classification with significant trends in 20.
(v) Assignee code controls are seven dummy variables that define who holds the rights to the patent. Table A.2 describes the assignee codes in further detail. The declining probability of field jump persists when the trend is examined within each assignment code, although the significance of the trend disappears in the rarer classifications.

Table 4.4: Trends in Time Lag

	Dependent Variable: Time Lag Between Consecutive Patent Applications					
	(1)	(2)	(3)	(4)	(5)	(6)
Application Year	0.30e-3 (.14e-3)	1.2e-3 (.14e-3)	0.54e-3 (.14e-3)	2.2e-3 (.35e-3)	2.8e-3 (.35e-3)	2.0e-3 (.35e-3)
Foreign Patent	--	-.0736 (.0016)	-.0591 (.0016)	--	-.0526 (.0042)	-.0522 (.0042)
Team Size (second patent)	--	-.0156 (.0004)	-.0099 (.0004)	--	--	--
Same Team Size Dummy	--	-.0474 (.0016)	-.0515 (.0016)	--	--	--
Field Jump Dummy	--	.115 (.002)	.115 (.002)	--	.081 (.004)	.083 (.004)
Technological Field Controls (second patent)	--	--	Yes	--	--	Yes
Number of observations	1,430,144	1,430,144	1,430,144	215,855	215,855	215,855
Period	1975-1993	1975-1993	1975-1993	1975-1993	1975-1993	1975-1993
Mean of Dependent Variable (see note (iv))	.749	.749	.749	.793	.793	.793
Per-decade Trend as % of Period Mean (see note (iv))	0.4%	1.6%	0.7%	2.8%	3.5%	2.5%
R^2	.0000	.0077	.0157	.0002	.0028	.0136

NOTES
(i) Regressions are OLS, with standard errors in parentheses.
(ii) All specifications consider “fast” innovators -- only those consecutive patents with no more than 3 years between applications and with no more than 3 years delay between application and grant. (See discussion in text.)
(iii) Specifications (1) to (3) consider all consecutive patents in this time period. Specifications (4) through (6) consider time lags by solo inventors.
(iv) The dependent variable is an integer varying between 0 and 3. Period means are underestimated due to the integer nature of the application year, because two applications in the same calendar year are calculated to have a time lag of zero. This biases down the mean and biases up the percentage trend.
(v) Technological field controls are dummies for the 36 categories defined by Hall et al (2001). Figure 4.3 presents the trend for each of these categories individually.

Table 4.5: Mean differences across Technological Categories

Technological Classification (Hall et al. 2001) 6	36	Code	Age at First Innovation Obs	Mean	Inventors per Patent Obs	Mean	Probability of Field Jump Obs	Mean
Chemical (1)	Agriculture, Food, Textiles	11	12	31.1	16,100	2.41	2,500	0.48
	Coating	12	53	29.2	29,800	2.23	4,300	0.64
	Gas	13	17	30.3	9,200	1.96	1,700	0.59
	Organic Compounds	14	51	29.5	59,600	2.56	7,000	0.34
	Resins	15	44	29.3	67,200	2.51	7,500	0.36
	Miscellaneous—Chemical	19	331	29.3	197,100	2.23	29,500	0.43
	Entire category		508	29.4	379,200	2.33	52,100	0.43
Computers & Communications (2)	Communications	21	264	29.3	92,700	1.99	15,000	0.41
	Computer Hardware & Software	22	162	29.8	80,400	2.26	10,200	0.44
	Computer Peripherals	23	37	29.3	22,100	2.37	2,800	0.51
	Information Storage	24	43	28.9	41,300	2.21	6,700	0.39
	Entire category		506	29.4	236,700	2.16	34,500	0.42
Drugs & Medical (3)	Drugs	31	74	29.9	65,200	2.90	6,300	0.25
	Surgery & Medical Instruments	32	268	29.8	59,900	1.86	12,400	0.29
	Biotechnology	33	46	30.5	22,700	2.75	1,800	0.38
	Misc—Drugs & Medical	39	68	29.1	13,600	1.66	3,500	0.35
	Entire category		456	29.8	161,500	2.39	23,800	0.29
Electrical & Electronic (4)	Electrical Devices	41	111	29.3	61,000	1.77	12,700	0.48
	Electrical Lighting	42	90	29.6	31,300	1.96	5,700	0.43
	Measuring & Testing	43	116	29.2	57,700	1.94	10,000	0.51
	Nuclear & X-rays	44	52	29.7	30,200	2.08	4,700	0.50
	Power Systems	45	128	29.4	68,900	1.94	13,000	0.51
	Semiconductor Devices	46	49	29.3	44,700	2.25	7,100	0.34
	Misc—Electrical	49	104	29.1	49,100	1.97	8,900	0.51
	Entire category		650	29.3	343,300	1.97	61,700	0.48
Mechanical (5)	Materials Processing & Handling	51	241	29.4	100,000	1.79	21,700	0.48
	Metal Working	52	87	28.8	58,100	2.11	10,400	0.54
	Motors, Engines & Parts	53	83	29.4	73,300	1.85	16,200	0.41
	Optics	54	57	29.0	48,000	2.15	8,100	0.37
	Transportation	55	273	29.0	56,800	1.66	12,000	0.45
	Misc—Mechanical	59	449	29.1	96,800	1.64	22,400	0.49
	Entire category		1,190	29.1	433,300	1.83	90,500	0.46
Others (6)	Agriculture, Husbandry, Food	61	250	29.1	41,200	1.75	7,600	0.41
	Amusement Devices	62	269	29.4	20,900	1.41	4,300	0.37
	Apparel & Textile	63	211	29.1	32,400	1.57	7,600	0.37
	Earth Working & Wells	64	100	29.6	27,800	1.69	6,600	0.36
	Furniture, House Fixtures	65	346	29.1	41,000	1.42	9,400	0.50
	Heating	66	58	30.0	26,300	1.75	6,100	0.48
	Pipes & Joints	67	45	29.2	17,100	1.58	4,500	0.61
	Receptacles	68	298	29.4	40,700	1.51	10,100	0.47
	Misc—Others	69	846	29.2	167,800	1.73	35,200	0.48
	Entire category		2,423	29.3	415,600	1.64	91,000	0.46

NOTES
(i) Age at first innovation includes observations of those innovators who appear after 1985 in the data set and between the ages of 23 and 33. Results are similar, with higher mean and even less variance, for 25-35 year olds.
(ii) Probability of field jump is probability of switching categories for solo innovators using 36-category measure.

Table 4.6: Inventors per Patent vs. Tree Size

	Dependent Variable: Inventors per Patent						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Normalized Variation in Tree Size	.0849 (.0010)	.0961 (.0010)	.0995 (.0011)	.120 (.001)	.133 (.001)	.107 (.001)	.152 (.001)
Normalized Variation in Tree Size, Squared	.0609 (.0007)	.0545 (.0007)	.0545 (.0007)	.0341 (.0007)	.0257 (.0009)	.0356 (.0011)	.0404 (.0009)
Foreign Patent	--	.446 (.002)	.442 (.002)	.420 (.002)	US Only	Foreign Only	.371 (.003)
Normalized Variation in Direct Citations Made	--	--	-.0094 (.0011)	--	--	--	--
Technological Field Controls	--	--	--	Yes	Yes	Yes	Yes
Application Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,969,908	1,969,908	1,969,908	1,969,908	1,103,402	866,506	1,330,210
Period	1975-1999	1975-1999	1975-1999	1975-1999	1975-1999	1975-1999	1985-1999
Mean of Dependent Variable	2.02	2.02	2.02	2.02	1.82	2.27	2.13
R^2	.026	.050	.050	.100	.090	.083	.079

NOTES
(i) Regressions are OLS with standard errors in parentheses. Specifications (1) through (4) consider the entire universe of patents applied for between 1975 and 1999. Specification (5) and (6) consider separately patents from domestic vs. foreign sources. Specification (7) considers cross-sections from the later part of the time period.
(ii) Normalized Variation in Tree Size is the deviation from the year mean tree size, divided by the year standard deviation in tree size. “Tree size” is the log of the number of nodes in the citations tree behind any patent.
(iii) Normalized Variation in Direct Citations Made captures variation in the number of citations to prior art listed on a patent application. It is the deviation from the year mean number of citations, divided by the year standard deviation in the number of citations.
(iv) Technological field controls include dummies for each of Hall et al.’s 36-category measure.
(v) The number of observations here is slightly smaller than for the time trend analysis in Table 4.1 because a few patents do not cite other US patents, hence no citation tree can be built; these patents are dropped from the analysis.

Table 4.7: Age vs. Tree Size

	Dependent Variable: Age at application for first patent							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normalized Variation in Tree Size	-.007 (.032)	-.005 (.036)	.114 (.035)	.084 (.040)	.059 (.043)	.097 (.030)	.113 (.046)	.030 (.026)
Team Size	--	-.054 (.027)	--	-.036 (.030)	-.038 (.030)	-.024 (.025)	.008 (.035)	-.029 (.019)
Normalized Variation in Direct Citations Made	--	--	--	--	.064 (.044)	--	--	--
Technological Field Controls	--	Yes	--	Yes	Yes	Yes	Yes	Yes
Application Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	6,486	6,486	5,058	5,058	5,058	8,434	3,630	3,588
Period	1985-1999	1985-1999	1985-1999	1985-1999	1985-1999	1975-1999	1985-1999	1985-1999
Age Range	25-35	25-35	23-33	23-33	23-33	23-33	21-31	28-33
Mean of Dependent Variable	31.0	31.0	29.34	29.3	29.2	29.2	27.7	30.7
R^2	.009	.022	.009	.021	.012	.020	.025	.020

NOTES

(i) Regressions are OLS, with standard errors in parentheses. All regressions look only at those innovators for whom we have age data. Specifications (1) and (2) consider first innovations in the 25-35 age window. Specifications (3) through (6) consider innovators in the 23-33 age window. Specification (7) considers slightly younger innovators, and Specification (8) considers the latter half of the 23-33 age window. Specifications (6) considers cross-sections pooled over the entire time period; the other specifications focus on the post-1985 period, for which we can be confident that we are witnessing an innovator's first patent.

(ii) Normalized Variation in Tree Size is the deviation from the year mean tree size, divided by the year standard deviation in tree size. "Tree size" is the log of the number of nodes in the citations tree behind any patent.

(iii) Normalized Variation in Direct Citations Made captures variation in the number of citations to prior art listed on a patent application. It is the deviation from the year mean number of citations, divided by the year standard deviation in the number of citations.

(iv) The number of observations here is slightly smaller than for the time trend analysis in Table 4.2 because a few patents do not cite other US patents, hence no citation tree can be built; these patents are dropped from the analysis.

(v) Technological field controls include dummies for each of Hall et al.'s 36-category measure.

Table 4.8: Field Jump vs. Tree Size

	Dependent Variable: Probability of Switching Technological Field					
	(1)	(2)	(3)	(4)	(5)	(6)
Normalized Variation in Tree Size	-.0072 (.0008)	-.0074 (.0008)	-.0059 (.0008)	-.0095 (.0009)	-.0144 (.0012)	-.0184 (.0017)
Foreign Patent	--	-.0125 (.0018)	-.0108 (.0018)	-.0129 (.0018)	-.0135 (.0023)	.0032 (.0032)
Time Between Applications	--	--	.0226 (.0004)	.0232 (.0004)	.0215 (.0012)	.0143 (.0017)
Technological Field Controls (first patent)	--	--	--	Yes	Yes	Yes
Application Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	353,762	353,762	353,762	353,762	212,274	110,511
Period	1975-1999	1975-1999	1975-1999	1975-1999	1975-1993	1985-1993
Mean of Dependent Variable	.551	.551	.551	.551	.536	.520
(Pseudo) R^2	.0039	.0039	.0117	.0251	.0171	.0159

NOTES
(i) Results are for probit estimation, with coefficients reported at mean values and z-statistics in parentheses. The coefficient for the Foreign dummy is reported over the 0-1 range. Only solo inventors are considered. Specifications (1) through (4) consider the entire set of solo inventors. Specification (5) considers only those solo inventors who meet the criteria in Specifications (1) through (6) in Table 4.3 (to help control for any truncation bias in the specialization measure – see the discussion of Table 4.3 in the text). Specification (6) considers the same data as Specification (5), but only looks at cross-sections in the later part of the time period.
(ii) The dependent variable is 0 if an inventor does not switch fields between two consecutive innovations. The field is defined using the 414-category technological class definition of the USPTO.
(iii) Normalized Variation in Tree Size is the deviation from the year mean tree size, divided by the year standard deviation in tree size. “Tree size” is the log of the number of nodes in the citations tree behind any patent.
(iv) Technological field controls include dummies for each of Hall et al.’s 36-category measure.

Table A.1: Number of Observations at Each Stage of Selection

	Number of Observations	Percentage of Row (3)	Percentage of Row (4)	Percentage of Row Above
(1) Patents Granted	2,139,313			
(2) Inventors Worldwide	4,301,229			
(3) Unique Inventors Worldwide	1,411,842			
(4) Unique Inventors with US Address	752,163	53.3%		53.3%
(5) Unique Inventors, US Address, Zip Code	224,152	15.9%	29.8%	29.8%
(6) Unique Inventors, US Address, Zip Code, Unique Match from AnyBirthday.com	56,281	4.0%	7.5%	25.1%

NOTES
(i) Observation counts consider the 1975-1999 period.
(ii) A "unique inventor" is defined by having same first name, last name, and middle initial.

Table A.2: The Assignment of Patent Rights

Assignment Status	All Patents	US Patents	US Patents No zip code	US Patents Zip code	Birth Data Direct Match	Birth Data Other Patents
Unassigned	17.2%	22.4%	0.4%	98.3%	97.9%	26.6%
US non-govt organization	43.9%	72.9%	94.1%	0.0%	0.0%	65.7%
Non-US non-govt organization	36.2%	1.1%	1.4%	0.0%	0.0%	3.4%
Other assignment	2.7%	3.5%	4.1%	1.7%	2.1%	4.4%

NOTES
(i) The first column considers all patent observations in the 1975-1999 period (2.1 million observations).
(ii) US patents are those for which first inventor listed with the patent has a US address.
(iii) The Birth Data columns consider those US patents with zip code information for which AnyBirthday.com produced a birth date. The first Birth Data column considers the specific patents on which AnyBirthday.com was able to match. The last column considers all other patents by that innovator, identifying the innovator by last name, first name, and middle initial.
(iv) Unassigned patents are those for which the patent rights were still held by the original inventor(s) at the time the patent was granted; these patents may or may not have been assigned after the grant date.
(v) Non-government organizations are mainly corporations but also include universities.
(vi) Other assignment includes assignments to: (a) US individuals; (b) Non-US individuals; (c) the US government; and (d) non-US governments.

Table A.3: Inventors per Patent, Mean Differences between Samples

	Dependent Variable: Inventors per patent				
	(1)	(2)	(3)	(4)	(5)
US Address dummy	-.315 (.0020)	-.339 (.0020)	-.300 (.0020)	-.124 (.0049)	-.103 (.0048)
US Address and Zip Code dummy	-.786 (.0033)	-.670 (.0033)	-.769 (.0032)	-.155 (.0069)	-.176 (.0066)
US Address, Zip Code, and AnyBirthday.com Direct Match dummy	.237 (.0068)	.246 (.0067)	.212 (.0067)	.243 (.0067)	.228 (.0066)
Constant	2.28 (.0014)	2.57 (.0023)	1.96 (.0052)	1.45 (.0042)	1.56 (.0067)
Technological Category dummies	No	Yes	No	No	Yes
Grant Year dummies	No	No	Yes	No	Yes
Assignee Code dummies	No	No	No	Yes	Yes
R^2	.0555	.0825	.0756	.0757	.1162

NOTES
(i) Regressions consider means in the entire dataset (2.1 million patent observations), covering the 1975-1999 time period. Standard errors are in parentheses.
(ii) Dummy variables are nested: The second row captures a subset of the first. The third row captures a subset of the second.
(iii) Innovators for whom AnyBirthday.com produces a birth date are often involved with multiple innovations over the 1975-1999 period. The patents used for comparison in this table are those patents for which AnyBirthday.com produced the direct match.
(iv) Regressions with technological category controls are reported using the 6-category measure of Hall et al (2001). Results using the 36-category measure are similar.

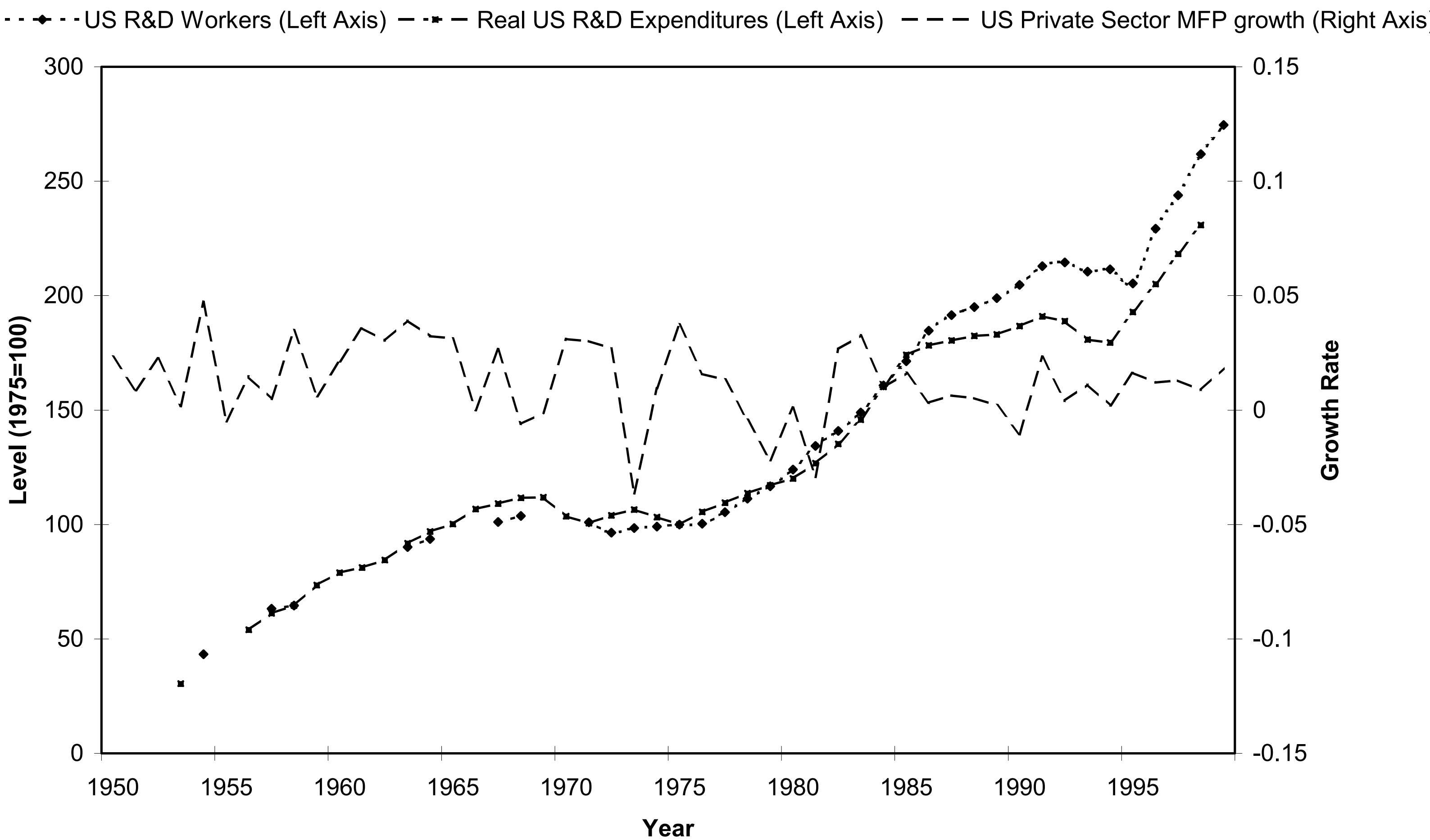

Figure 2.1: Rising Research Intensity, Flat Productivity Growth

Figure 2.2: Trends per R&D Worker in R&D Expenditures and Patent Grants

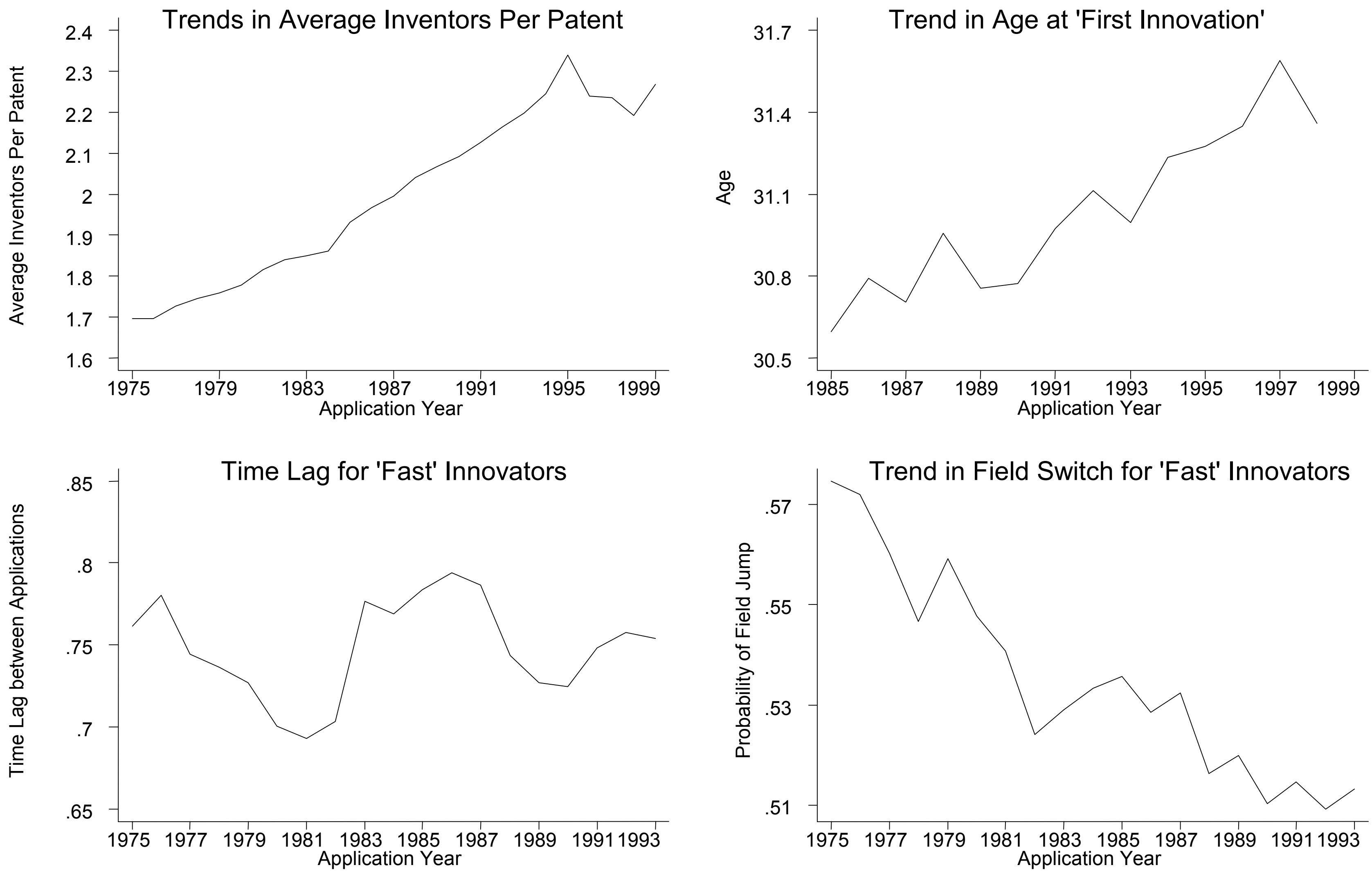

Figure 4.2: Basic Time Trends

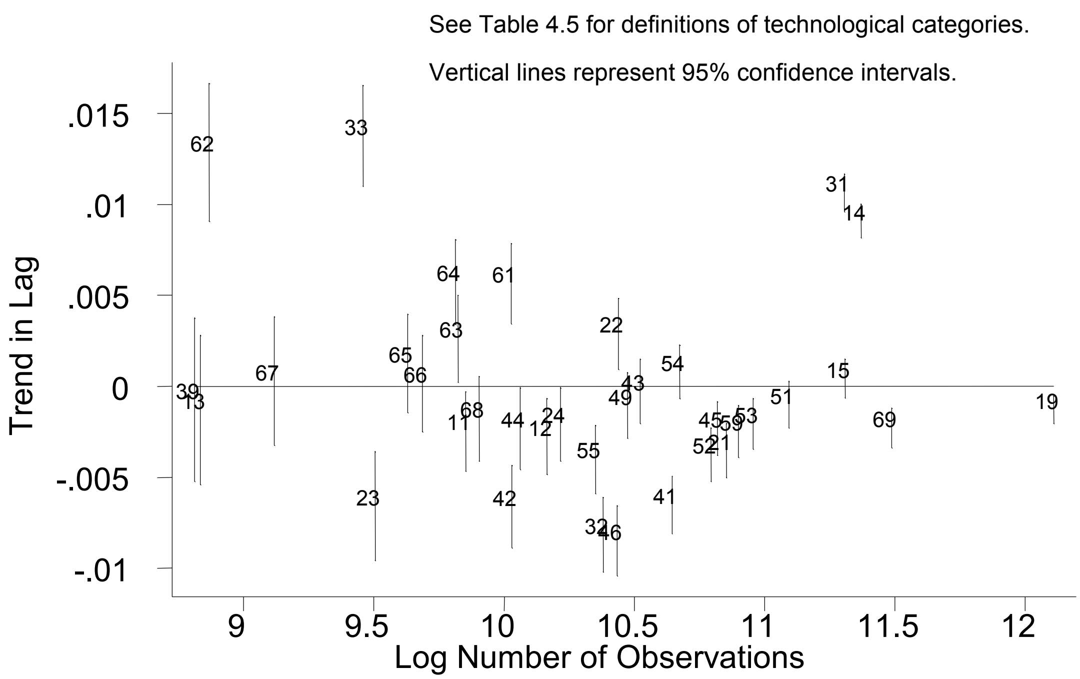

Figure 4.3: Trends in Time Lag by Technological Category

Epanichnikov kernel, bandwidth=0.2

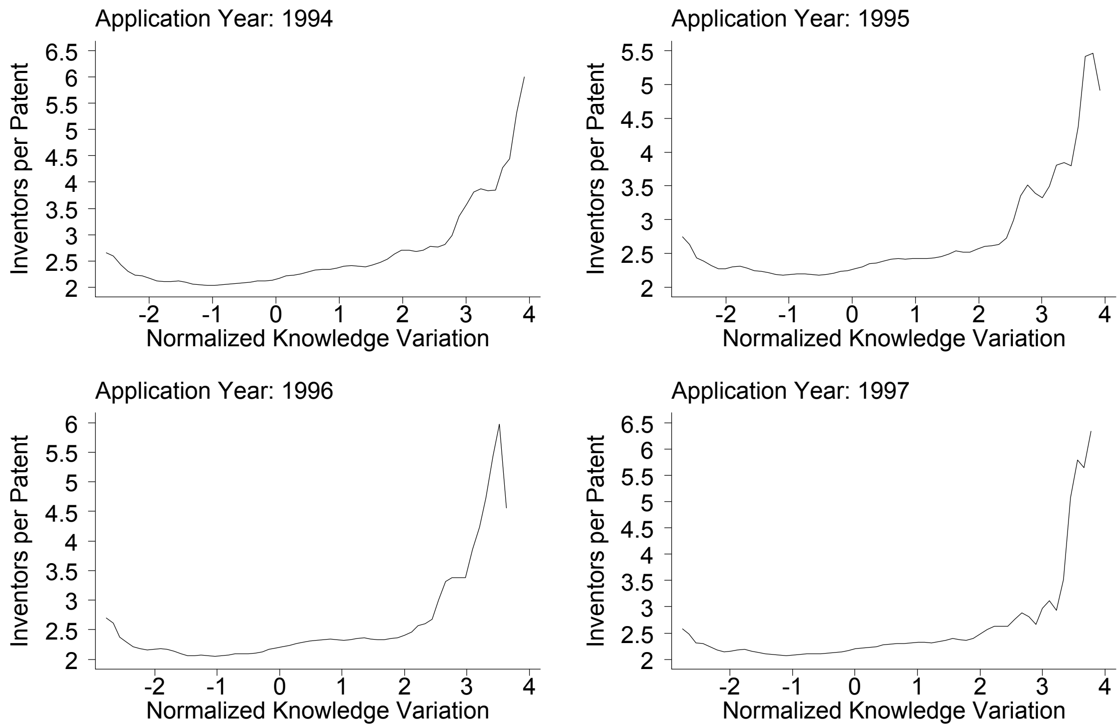

Figure 4.4: Team Size vs. Knowledge Measure, by Application Year

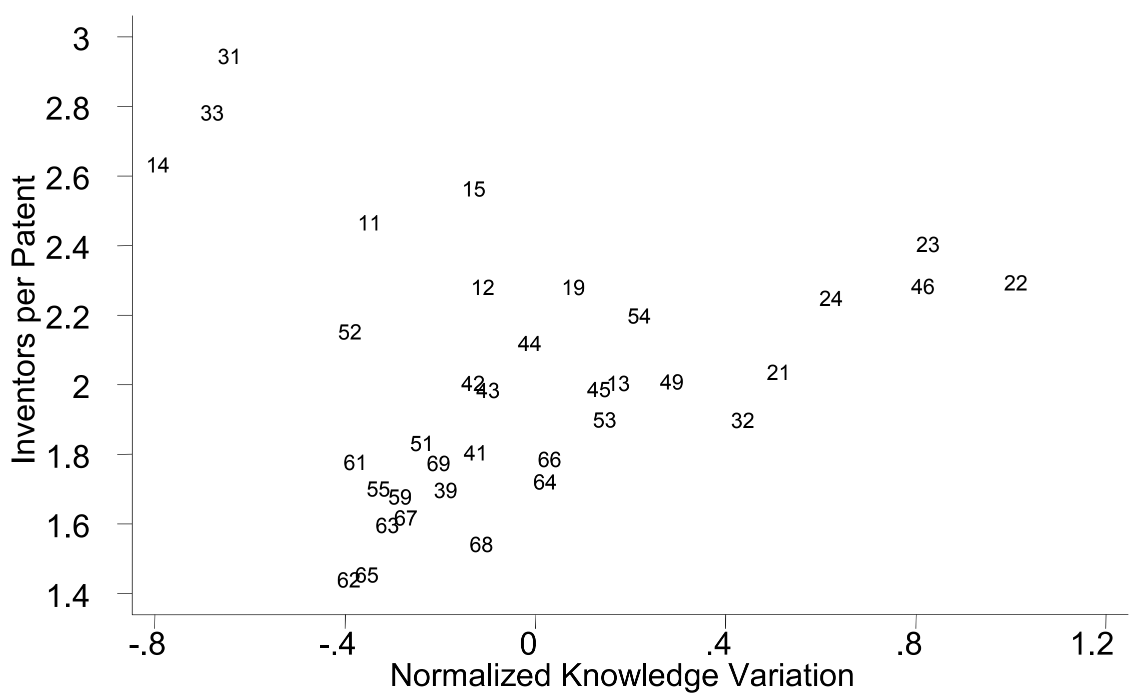

Figure 4.5: Team Size vs. Tree Size, Averages by Technological Category