Essays on Innovation, Leadership, and Growth

by

Benjamin F. Jones

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2003

Signature of Author ....
Department of Economics
August 15, 2003

Certified by ....
Daron Acemoglu
Professor of Economics
Thesis Supervisor

Certified by ....
Abhijit Banerjee
Professor of Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Elisha Gray II Professor of Economics
Chairperson, Department Committee on Graduate Students

Essays on Innovation, Leadership, and Growth

by

Benjamin F. Jones

Submitted to the Department of Economics
on August 15, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

The first two chapters of this thesis investigate a possibly fundamental aspect of technological progress. If knowledge accumulates as technology progresses, then successive generations of innovators may face an increasing educational burden, forcing them to spend longer periods in education and/or become increasingly specialized. In either case, the productivity of innovators may be reduced, with negative implications for growth.

The first chapter develops a formal model to examine the growth implications of this "knowledge burden mechanism" and generate testable predictions for innovators. The model predicts that educational attainment, specialization, and teamwork will rise over time. 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 second chapter further investigates the knowledge burden mechanism. Using data on famous inventions, I find that the age at which inventors produced them increased by 5 years over the 20th Century. The chapter employs a maximum likelihood model to test explanations for this trend. A knowledge-based explanation receives considerable support: innovators appear to arrive at the knowledge frontier 6.6 years later at the end of the 20th Century than they did at the beginning. This trend is unlikely to be sustainable and further suggests that educational externalities are a problematic byproduct of technological progress, particularly if innovators do their best work when they are young.

The final chapter investigates whether national leaders impact growth in developing countries. Using deaths of leaders while in office as a source of exogenous variation, we ask whether such randomly-timed leadership transitions are associated with shifts in country growth rates. We find robust evidence that leaders have a causal effect on growth. The effect of leaders on growth appears limited to non-democracies, where the death of an autocrat tends to be followed by a substantial, prolonged increase in growth.

Thesis Supervisor: Daron Acemoglu
Title: Professor of Economics

Thesis Supervisor: Abhijit Banerjee
Title: Professor of Economics

Contents

2 The Burden of Knowledge: Evidence from Great Inventions 70

3 Do Leaders Matter? National Leadership and Growth in the Developing World[1] 98

---

[1] This research has been performed jointly with Ben Olken.

Acknowledgments

There are many people I would like to thank for their advice and guidance.

Daron Acemoglu has provided invaluable advice at all levels, from broad discussions of research directions to close analysis of models and econometric work. He has encouraged my research interests while forcing me to bring rigor to them, and I believe I have learned my most valuable lessons in economics at his hands. His door has always been open, and I am grateful for and inspired by his generosity of intellect and time.

Abhijit Banerjee has also provided extraordinarily helpful insights on my research and needed encouragement. His words have, at important junctures, caused me to greatly rethink my approaches, and my thesis has been much improved by his advice. He has also been a willing ear throughout my studies at MIT, even when my research methods have lagged (as they often still do) behind the questions I wish to investigate.

I would also like to especially thank Sendhil Mullainathan and Esther Duflo for their guidance and incisive advice. Ben Olken, my co-author for the third chapter, has been wonderful to work with, and Amy Finkelstein has also been a constant source of advice throughout my studies. I have been lucky to spend time in this great community of economists, and I could thank so many further members of the department, as well as my peers, that it is difficult to name any further without doing an injustice to some.

Financial support from the MIT Department of Economics, the George Shultz Fund, and the Program in Applied Economics of the Social Science Research Council, with funding by the John D. and Catherine T. MacArthur Foundation, is gratefully acknowledged.

Finally, I would like to thank my wife, Juliet, and my family for their unflagging support, especially in this last year when time has been increasingly scarce.

I dedicate this thesis to my father, Michael E. Jones, with the hope that he would have found the following chapters worth discussion and the endeavor of this doctorate a worthwhile pursuit.

Chapter 1

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

1.1 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 chapter 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 chapter.

This chapter 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 chapter 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.

1.2 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 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

---

[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.

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 chapter 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.

1.3 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 income by licensing their

---

[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.

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.

1.3.1 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,

$$y_j(t) = X_j(t) l_j(t) \tag{1.1}$$

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:

$$Y(t) = X(t)L_Y(t) \tag{1.2}$$

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:

$$w(t) = X(t) - r(t) \tag{1.3}$$

1.3.2 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]

$$U(\tau) = \int_{\tau}^{\infty} c(t)e^{-\phi(t-\tau)}dt \tag{1.4}$$

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.

---

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

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:

$$U^{wage}(\tau) = \int_{\tau}^{\infty} w(t)e^{-\phi(t-\tau)}dt \tag{1.5}$$

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,

$$U_i^{R\&D}(\tau) = \int_{\tau}^{\infty} v_i(\tau,t)e^{-\phi(t-\tau)}dt - E_i(\tau) \tag{1.6}$$

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

1.3.3 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

---

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

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:

$$E_i(\tau) = (b_i D(\tau))^{\varepsilon} \tag{1.7}$$

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.

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,

$$D(t) = (X(t))^{\delta} \tag{1.8}$$

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).

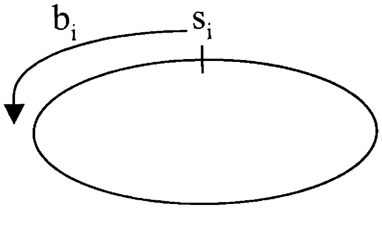


Figure 3.1: The circle of knowledge

1.3.4 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 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.

---

[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.

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]

$$V = \gamma \int_{t}^{t+z} L_Y(\widetilde{t})d\widetilde{t} \tag{1.9}$$

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 (1.9) to find: $V = \gamma C L_Y$, where $C = \left(e^{g_L z} - 1\right)/g_L$, 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

---

[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.

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,

$$\begin{aligned} \lambda_i(\tau, t) &= X(\tau)^{\chi_\lambda} L(t, s_i)^{-\sigma} b_i^{\beta_\lambda} && (1.10) \\ \gamma_i(\tau, t) &= X(\tau)^{\chi_\gamma} b_i^{\beta_\gamma} && (1.11) \end{aligned}$$

and therefore

$$\theta_i(\tau, t) = X(\tau)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} \tag{1.12}$$

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 (1.10) and (1.11), I can now explicitly define an innovator's expected flow of income,

$$v_i(\tau, t) = X(\tau)^\chi L(t, s_i)^{-\sigma} b_i^\beta C L_Y(t) \qquad (1.13)$$

---

[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 (1.12) 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.

1.3.5 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 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:

$$\begin{aligned} U_i^{R\&D}(s_i, b_i) &\geq U_i^{R\&D}(s, b) \quad \forall s, b \;\; \forall i \in l_R(\tau) && (1.14) \\ U_i^{R\&D}(s_i, b_i) &= U^{wage} \qquad\quad \forall i \in l_R(\tau) \;\; \forall j \in l_Y(\tau) && (1.15) \end{aligned}$$

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 (1.14) and (1.15), the equilibrium outcome.

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

---

[13] Condition (1.15) 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 (1.15).

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.

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 (1.5) to find:

$$U^{wage}(\tau) = \frac{X(\tau)}{\phi - g} e^{-gz} \tag{1.16}$$

Innovators

Using (1.6) and the equilibrium condition (1.14), the innovator's problem is:

$$\max_{s_i, b_i} \int_{\tau}^{\infty} v_i(\tau, t) e^{-\phi(t-\tau)} dt - E_i(\tau) \tag{1.17}$$

Or, with the definitions of $E_i(\tau)$ and $v_i(\tau, t)$ in (1.7) and (1.13),

$$\max_{s_i, b_i} \int_{\tau}^{\infty} X(\tau)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} C L_Y(t) e^{-\phi(t-\tau)} dt - (b_i D(\tau))^{\varepsilon} \tag{1.18}$$

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 (1.18) with respect to b_i produces the first order condition:

$$\frac{\beta}{b_i^*} C X(\tau)^{\chi} b_i^{*\beta} \int_{\tau}^{\infty} L_R(t)^{-\sigma} L_Y(t) e^{-\phi(t-\tau)} dt = \frac{\varepsilon}{b_i^*} \left(b_i^* D(\tau)\right)^{\varepsilon} \tag{1.19}$$

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,

$$b^*(\tau) = \frac{1}{D(\tau)} \left(\frac{v(\tau)}{\phi - (1-\sigma) g_L} \frac{\beta}{\varepsilon} \right)^{1/\varepsilon} \tag{1.20}$$

where I write $b^*(\tau)$ to acknowledge that innovators in the same cohort face the same maximiza-

---

[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 (1.19) 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.

tion 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,

$$E^*(\tau) = \frac{v(\tau)}{\phi - (1-\sigma)g_L}\frac{\beta}{\varepsilon} \tag{1.21}$$

$$U^{R\&D*}(\tau) = \frac{v(\tau)}{\phi - (1-\sigma)g_L}\left(1 - \frac{\beta}{\varepsilon}\right) \tag{1.22}$$

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 (1.15). 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.

1.3.6 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 $\overline{\theta}(t)$, then productivity increases at rate $dX/dt = \overline{\theta}(t)L_R(t)$. The growth rate in the economy is then,

$$g = \frac{\overline{\theta}(t)L_R(t)}{X(t)} \tag{1.23}$$

---

[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 (1.20) and rearranging shows that $b^*(\tau) = \left(\frac{1}{D(\tau)}\right)^{\varepsilon/(\varepsilon-\beta)} \left(\frac{CX(\tau)^{\chi}L_R(\tau)^{-\sigma}L_Y(\tau)}{\phi-(1-\sigma)g_L}\frac{\beta}{\varepsilon}\right)^{1/(\varepsilon-\beta)}$.

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,

$$g = \frac{1-\sigma}{1-\chi-\beta(\frac{1}{\varepsilon}-\delta)} g_L \tag{1.24}$$

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 (1.24). 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 (1.24),

$$g_{b^*} = \left(\frac{1}{\varepsilon} - \delta\right) g \tag{1.25}$$

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 (1.24) 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 chapter, 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 (1.24) 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]

---

[16] With $g_L = 0$, we see from (1.23) that steady-state growth requires $g = g_{\bar{\theta}}$. As can be seen from (1.42) in the Appendix and (1.25), 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).

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

1.3.7 Time Series Predictions

In addition to its predictions for the evolution of specialization (equation (1.25)), 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 (1.21)) 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.

$$g_{E^*}=g \tag{1.26}$$

This is seen formally by taking logs in (1.21), 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]

$$\overline{team}(b) = n - \sum_{j=1}^{n-1} F(jb) \tag{1.27}$$

where j indexes a particular realization of team size. Differentiating (1.27) with respect to b shows that,

---

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

$$\frac{d\overline{team}(b)}{db} = -\sum_{j=1}^{n-1} jf(jb) \tag{1.28}$$

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]

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]

$$\overline{lag} = \frac{\int_t^\infty h\tilde{t}e^{-(h+\phi)(\tilde{t}-t)}d\tilde{t}}{\int_t^\infty he^{-(h+\phi)(\tilde{t}-t)}d\tilde{t}} = \frac{1}{\lambda\overline{team}(b) + \phi} \tag{1.29}$$

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 –

---

[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.

[20] Given that you have innovated at time t, the probability that you neither innovate again nor die by time $\tilde{t}$ is $e^{-(h+\phi)(\tilde{t}-t)}$, and the hazard rate of innovating at any time $\tilde{t}$ is h. The numerator of (1.29) 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).

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.

1.3.8 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 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:

$$E_{ij}(\tau) = (b_i D_j(\tau))^{\varepsilon} \tag{1.30}$$

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

$$\theta_{ij}(t,\tau) = A_j(t) X(\tau)^{\chi} L_j(t, s_{ij})^{-\sigma} b_{ij}^{\beta} \tag{1.31}$$

This specification differs in two ways from that in equation (1.12). 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:

$$\frac{\beta}{b_{ij}^*} C X(\tau)^{\chi} b_i^{*\beta} \int_{\tau}^{\infty} A_j(t) L_{Rj}(t)^{-\sigma} L_Y(t) e^{-\phi(t-\tau)} dt = \frac{\varepsilon}{b_{ij}^*} \left(b_{ij}^* D_j(\tau)\right)^{\varepsilon} \tag{1.32}$$

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

$$b_j^*(\tau) = \frac{1}{D_j(\tau)} \left(\frac{v_j(\tau)}{\phi - (1-\sigma) g_L - g_{Aj}} \frac{\beta}{\varepsilon} \right)^{1/\varepsilon} \tag{1.33}$$

$$E_j^*(\tau) = \frac{v_j(\tau)}{\phi - (1-\sigma) g_L - g_{Aj}} \frac{\beta}{\varepsilon} \tag{1.34}$$

$$U_j^*(\tau) = \frac{v_j(\tau)}{\phi - (1-\sigma) g_L - g_{Aj}} \left(1 - \frac{\beta}{\varepsilon}\right) \tag{1.35}$$

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 (1.33) we see directly that,

$$\frac{b_j^*(\tau)}{b_{j'}^*(\tau)} = \frac{D_{j'}(\tau)}{D_j(\tau)} \tag{1.36}$$

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.

1.4 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.

1.4.1 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.

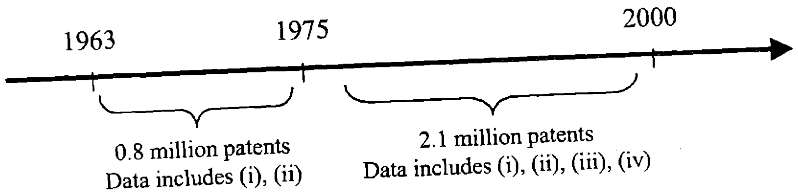


Figure 4.1: Summary of Available Data

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.

1.4.2 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" innova-

---

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

tion 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 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.

---

[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.

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 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.

---

[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 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.

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 subsample 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 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.

1.4.3 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",

---

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

"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 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

---

[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.

[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

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]

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 technologi-

---

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), 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).

cal 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 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.

1.4.4 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 chapter.

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 is that the underlying arrival rate of ideas, λ, may be declining. From equation (1.10), we see three possible explanations for

---

[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.

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.

1.5 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 chapter 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 chapter 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.

1.6 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 (1.15), 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 (1.6) that allows for the possibility that an idea is infeasible:

$$CX(t')^\chi b_i^{*\beta} \int_{t'}^{\infty} F(p_i(t))L(t, s')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt - \left(b_i^* D(t')\right)^\varepsilon \qquad (1.37)$$

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 (1.37) and the equilibrium condition (1.14), the choice s' can only be an equilibrium for person i if

---

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

$$\int_{t'}^{\infty} L(t,s')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt \geq \int_{t'}^{\infty} L(t,s'')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt \tag{1.38}$$

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

$$\int_{t''}^{\infty} L(t,s')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt > \int_{t''}^{\infty} L(t,s'')^{-\sigma} L_Y(t) e^{-\phi(t-t')} dt \tag{1.39}$$

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 (1.39) holds. Hence (1.38) 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 (1.38), and by extension (1.39). But there is no t'' at which (1.39) 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 (1.23), the steady-state growth rate in the economy is defined by,

$$g = g_{\overline{\theta}} + g_L \tag{1.40}$$

To define $g_{\overline{\theta}}$, 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]

$$\overline{\theta}(t) = \int_{-\infty}^{t} \theta(\tau, t)(g_L + \phi) e^{(g_L+\phi)(\tau - t)} d\tau \tag{1.41}$$

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 (1.12)), using the equilibrium result $b_i = b^*(\tau)$, and differentiating with respect to τ. We can therefore integrate (1.41) 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 $g = g_{\overline{\theta}} + g_L = g_\theta + g_L$.

$\theta(t)$ 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,

---

[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 $(g_L + \phi)e^{-g_L(t-\tau)}e^{-\phi(t-\tau)}$.

$$g_{\bar{\theta}} = \chi g + \beta g_{b^*} - \sigma g_L \tag{1.42}$$

Taking logs in the equilibrium result (1.20), 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 (1.13) that $v(t) = \theta(t)CL_Y(t)$, the growth rate in v is the same as the growth rate in the economy: $g_v = g_\theta + g_L = g$. From (1.8), the growth rate in the depth of knowledge is $g_D = \delta g$. Therefore,

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

Inserting this into (1.42), the result into (1.40), and rearranging produces the expression for steady-state growth in equation (1.24).

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 (1.27),

$$\overline{team}(b) = n - \sum_{j=1}^{n-1} F(jb;b) \tag{1.43}$$

Noting that $F(jb;b) = \int_0^{jb} f(k;b)dk$, we can use Leibniz's rule to differentiate (1.43) with respect to b and thereby define a necessary and sufficient condition for team size to increase with specialization:

$$\sum_{j=1}^{n-1} \left(\int_0^{jb} \frac{df(k;b)}{db} dk + jf(jb;b) \right) > 0 \Leftrightarrow \frac{d\overline{team}(b)}{db} < 0 \tag{1.44}$$

The second term on the left hand side is recognized from equation (1.28) 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 (1.44), 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.

1.7 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

---

[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.

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 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.

Bibliography

[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	.0261	.0262	.0251	.0244	.0306	.0180
		(.0001)	(.0001)	(.0001)	(.0001)	(.0002)	(.0002)	(.0003)
Foreign Patent		--	.444	.416	.141	.146	US Only	Foreign Only
			(.002)	(.002)	(.004)	(.004)		
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	.0666	.0671	.0671	.0687	.0530	.0584
		(.0095)	(.0095)	(.0095)	(.0099)	(.0097)	(.0107)	(.0109)
Technological Field Controls	Broad	--	Yes	--	--	--	--	--
	Narrow	--	--	Yes	Yes	Yes	--	Yes
Assignee Code		--	--	--	Yes	Yes	--	Yes
Team Size		--	--	--	--	-.0630	--	-.0348
						(.0273)		(.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)			Age at First Innovation		Inventors per Patent		Probability of Field Jump	
6	36	Code	Obs	Mean	Obs	Mean	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.

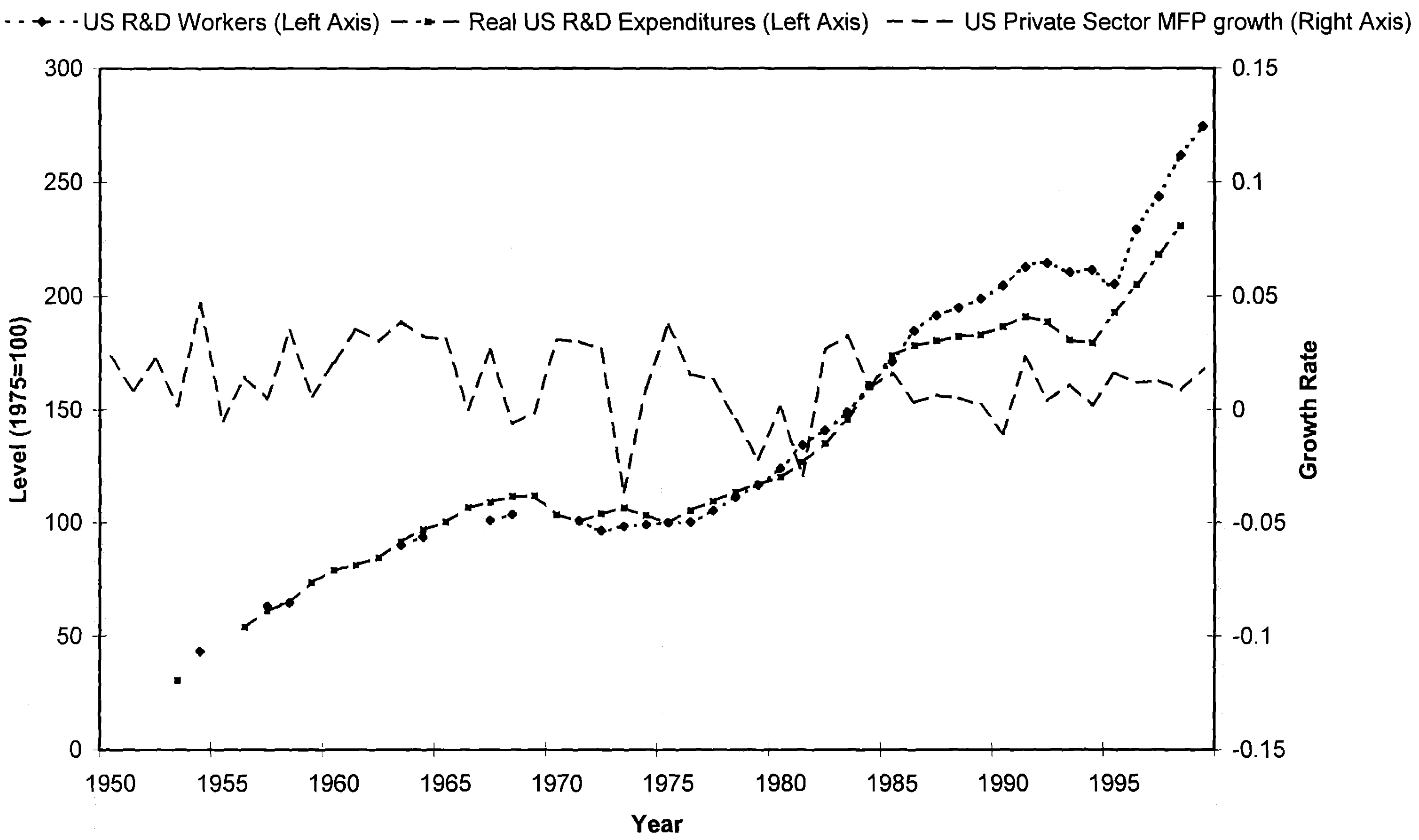


Figure 2.1: Rising Research Intensity, Flat Productivity Growth

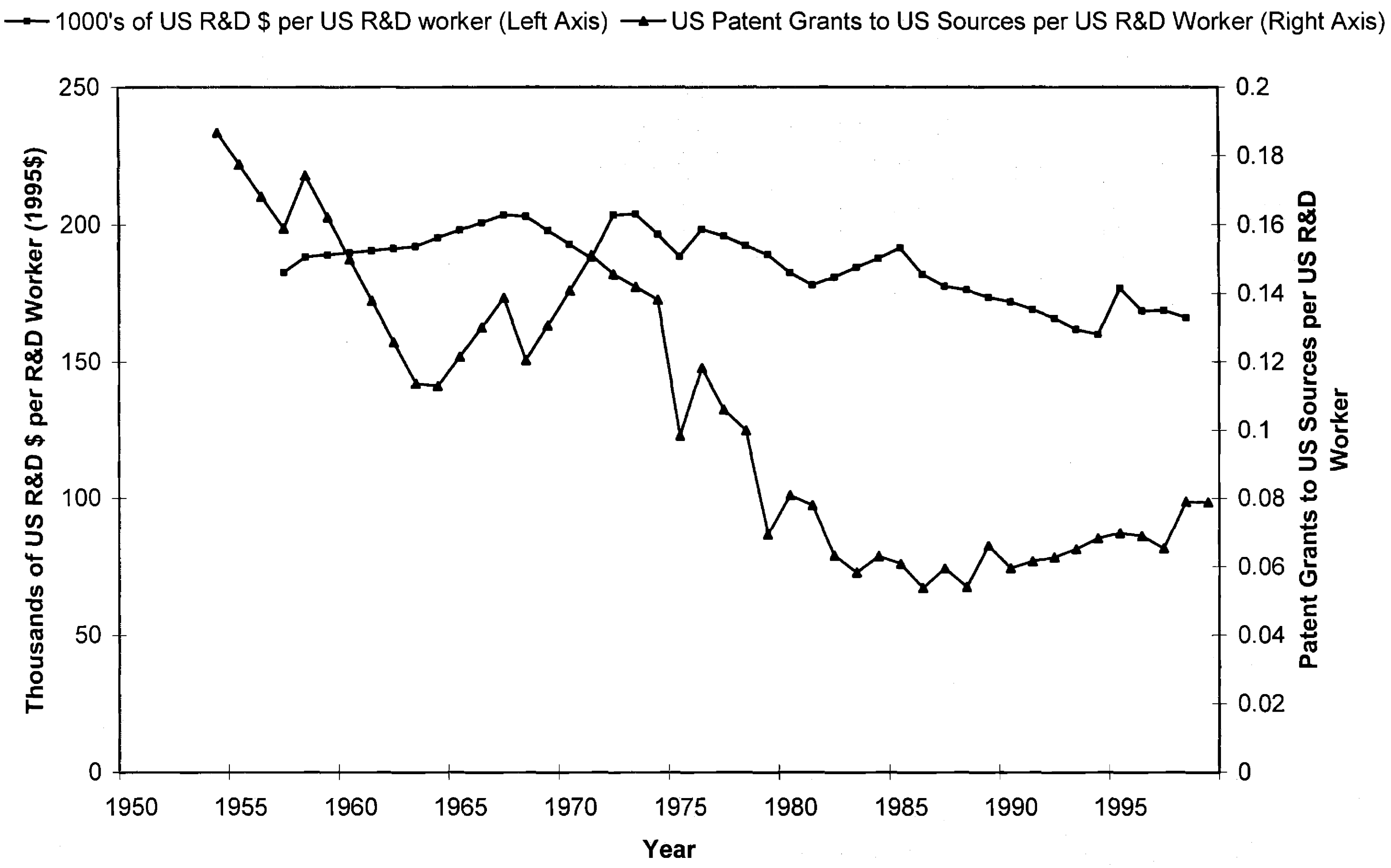


Figure 2.2: Trends per R&D Worker in R&D Expenditures and Patent Grants

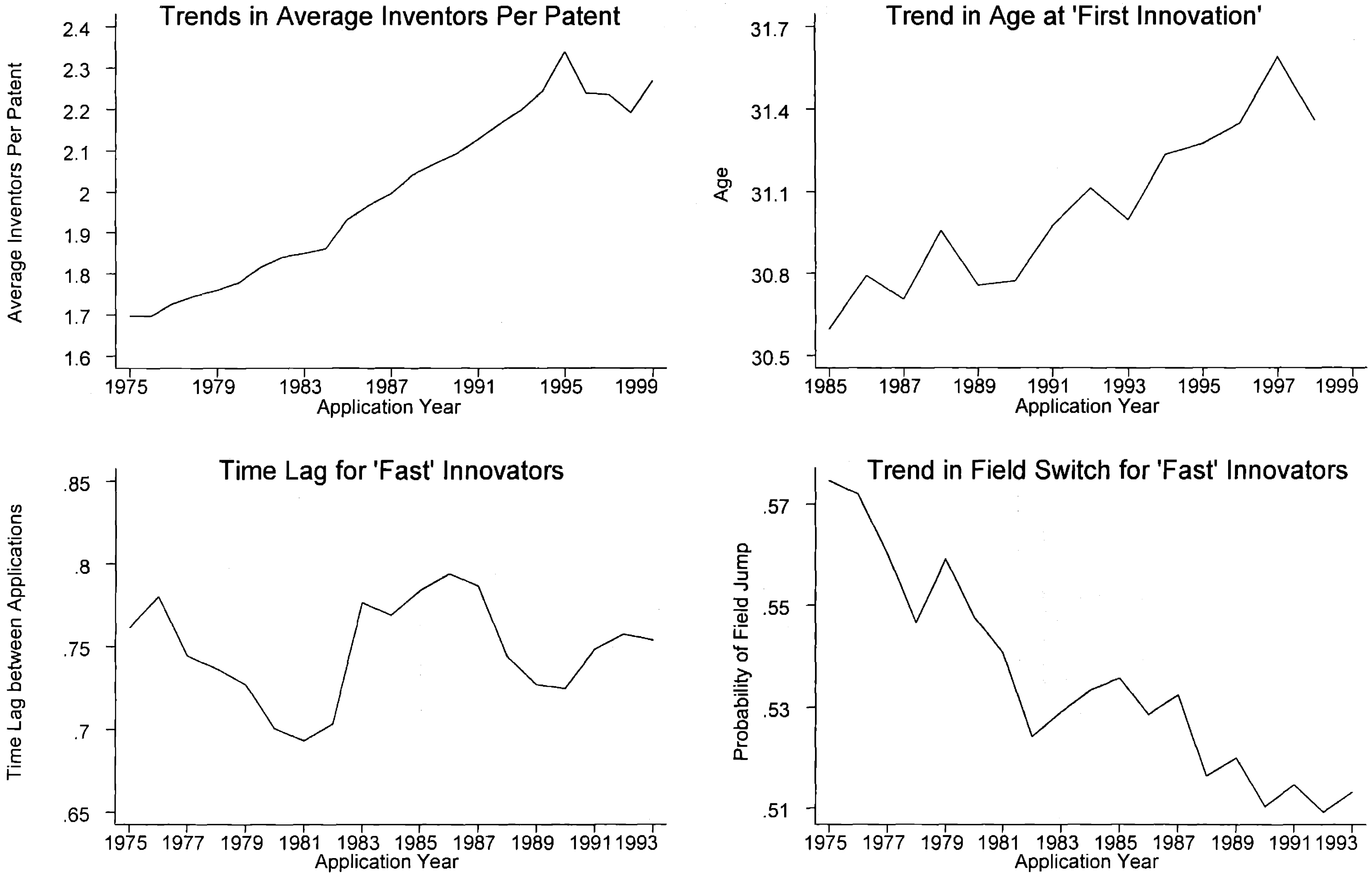
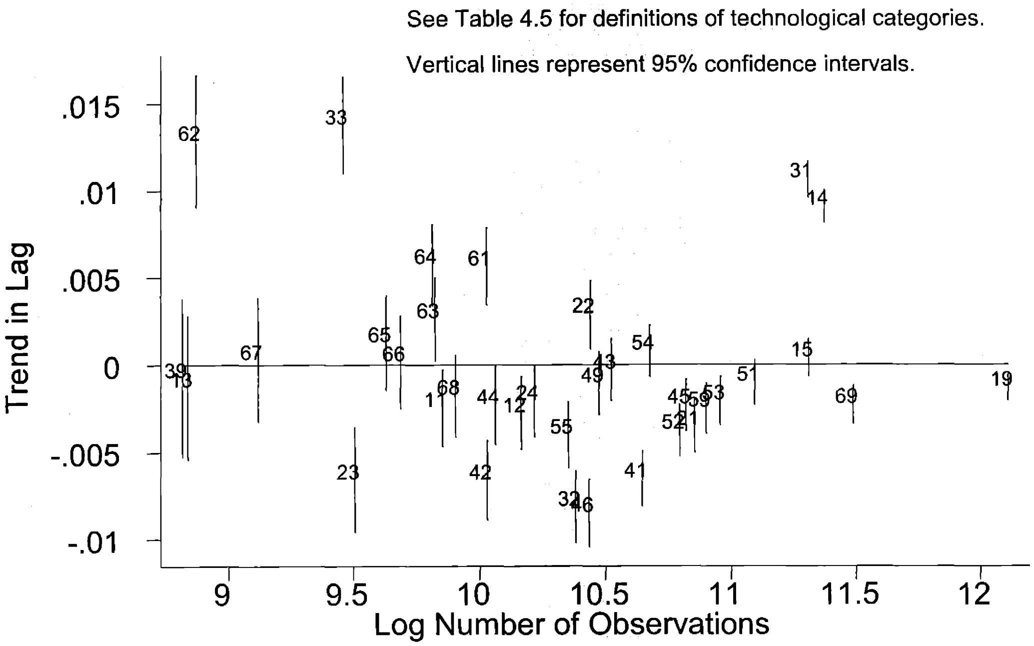


Figure 4.2: Basic Time Trends


See Table 4.5 for definitions of technological categories.

Vertical lines represent 95% confidence intervals.

Figure 4.3: Trends in Time Lag by Technological Category

Epanichnikov kernel, bandwidth=0.2

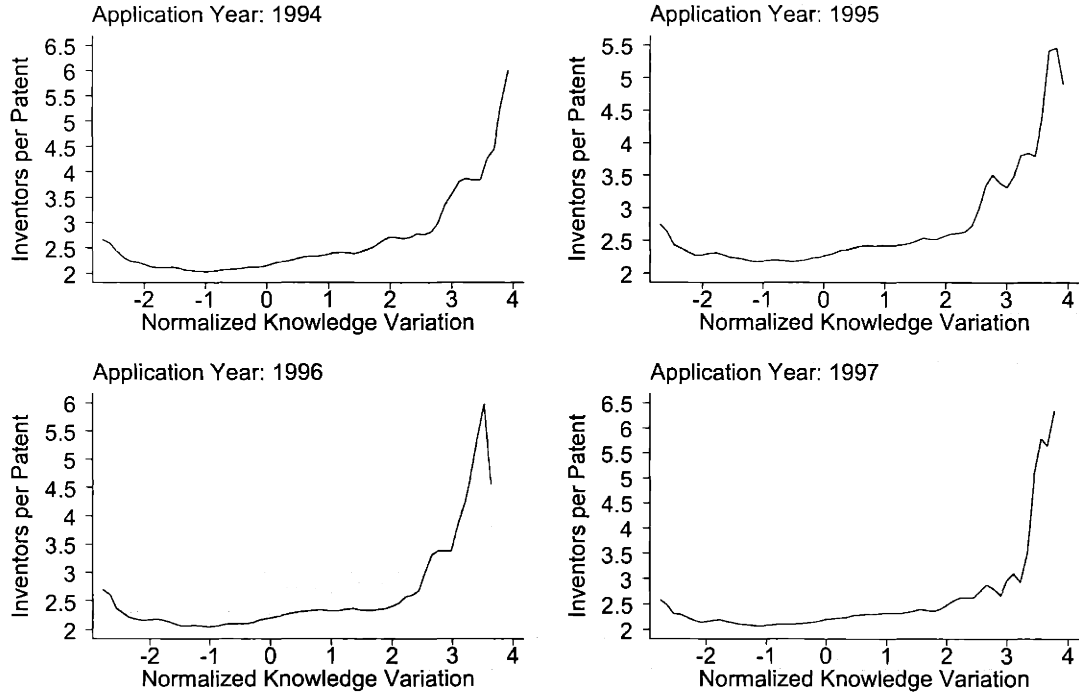


Figure 4.4: Team Size vs. Knowledge Measure, by Application Year

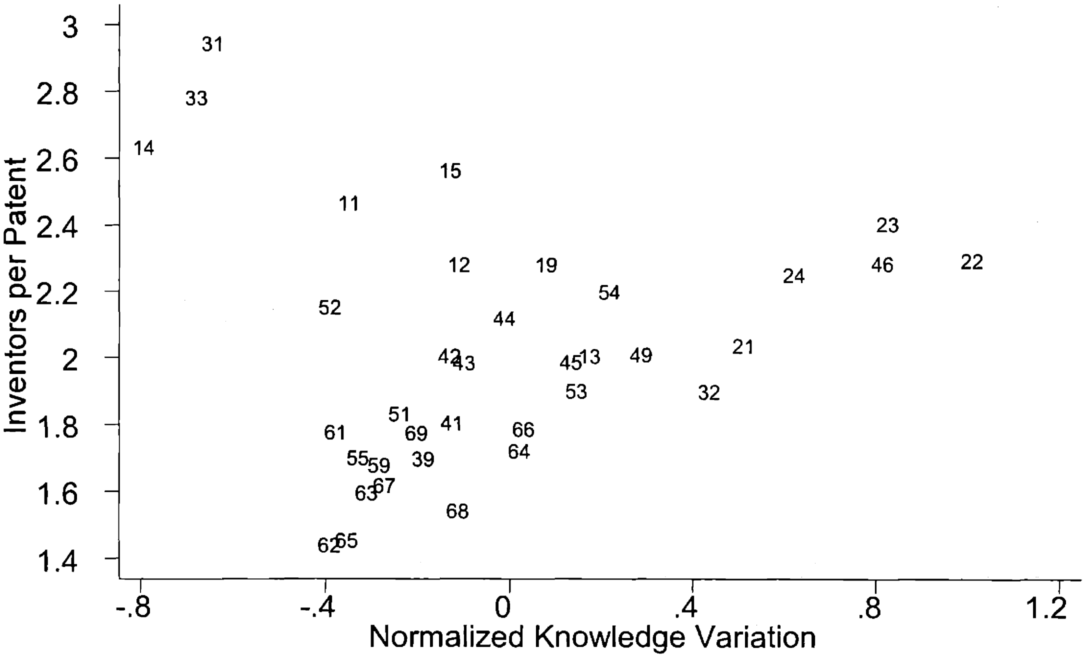


Figure 4.5: Team Size vs. Tree Size, Averages by Technological Category

Chapter 2

The Burden of Knowledge: Evidence from Great Inventions

Age is, of course, a fever chill
that every physicist must fear.
He's better dead than living still
when once he's past his thirtieth year.

– Paul Dirac, 1933 Nobel Laureate in Physics

2.1 Introduction

It is widely perceived that great innovations are the provenance of the young. The sentiments of Dirac expressed above have been shared by Einstein and many other eminent scientists and mathematicians (Simonton, 1988; Guterman, 2000). Empirical investigations of this view, usually undertaken within the fields of psychology and sociology, have made explicit measurements of the life-cycle output of innovators. This research tends to refute the idea that innovators necessarily produce their best work by the age of thirty. It does however support the idea that innovative activity is considerably greater at younger ages; in particular, innovative activity rises steeply in the 20's and 30's, peaks in the late 30's or early 40's, and then trails off through later years (Lehman, 1953; Simonton, 1984).

Perhaps the most intriguing empirical finding in Chapter 1 of this thesis is that the age at first innovation is rising over time. If this upward age trend is, as suggested in Chapter 1, an outcome of a rising burden of knowledge, then we can immediately see a troublesome interaction between rising learning requirements and the sentiment expressed by Dirac. Clearly, given a fixed retirement age, the longer an individual spends in education, the less time they will have left over for innovation (or any other productive activity). If we consider the early years of the life-cycle to be the most productive, as Dirac would do and the empirical literature suggests, then increasing educational requirements become that much more costly.

The model in Chapter 1 allowed for increasing educational attainment, but the time dimension of education was ignored. For tractability the model considered education as a lump-sum investment made immediately at birth. In consequence, any impact of a rising burden of knowledge on the productivity of innovators came entirely through the narrowing of innovators' expertise. A more general model can introduce the time dimension of education, with two effects. First, this introduces income opportunity costs into the educational decisions of innovators, which will alter their educational choices. Second, this allows the model to explicitly investigate the implications of lengthening education for the lifetime productivity of innovators, with further implications for growth. In a model not presented in this thesis (but available from the author), the time dimension of education is explicitly introduced. An interesting result from this exercise is that, should innovators internalize the entire benefit of their own education, they will *not* increase their educational attainment in response to a rising burden of knowledge. The intuition for this result follows from the time dimension of education: as the economy grows, wages rise at the same rate. Along a balanced growth path, the innovator faces a rising marginal cost (in foregone wages) to increasing their education that exactly offsets the marginal benefits of increasing their education. The equilibrium educational attainment of education is therefore fixed across innovator cohorts, and a rising knowledge burden is met entirely through increased specialization. In contrast, in the more realistic case where innovators share income from their innovations, the extended model shows that innovators will respond to an increased knowledge burden by increasing the length of their education. While this result, with its more realistic rent-sharing behavior, reestablishes the prediction found in Chapter 1, these modeling extensions are summarized here to suggest that the effects of an increasing knowledge burden on the age at first innovation is not entirely certain theoretically.[1]

The empirical evidence on the age at first innovation is also subject to some caveats. First, the age analysis in Chapter 1 is limited to a 15-year period at the end of the 20th Century. Interpreting trends of any historical importance from such a short time span is dubious. Second, unlike the other data in Chapter 1, the age data covers only a small subset of the universe of recent U.S. patent-holders. While concerns over selection bias in the age data were investigated rigorously, some concerns over selection may remain. A third issue, both empirical and theoretical in nature, is whether we should focus on age trends in the innovative population at large (what was studied in Chapter 1) or whether we should focus on a subset of very important inventors. In one view, the majority of innovations are unimportant and we should only be concerned with the best of the best. Put a slightly different way, any age trend among the general population of innovators may not extend to Einstein, Dirac, Bill Gates, or others of particular interest.[2]

This chapter presents new evidence indicating that innovators are reaching the knowledge frontier at later ages over time. To address some of the concerns listed above, the empirical work focuses on "great inventors" over a much longer time horizon: the entire 20th Century. The data considered includes both Nobel Prize winners in physics and inventors responsible for the 20th Century's great inventions, taken from technological almanacs.

---

[1] As a general matter, apart from any specific model formulation, it is clear that innovators can react to knowledge requirements on multiple dimensions, including both their degree of specialization and their time spent in education. While it may be most natural to think that inventors respond to an increasing knowledge burden partly through increasing specialization and partly through increasing educational attainment, the multiple dimensions of innovators' choice will allow some models to produce ambiguous theoretical predictions on any particular dimension of choice.

[2] Concerns about quality heterogeneity can be partly mitigated within the empirical framework of Chapter 1. Using a result from Trajtenberg, who found that the number of citations received by a patent correlates with its social value (Trajtenberg 1990), we can use the number of citations received as controls in the age regressions in Chapter 1. Such controls for quality have no effect on the age trend estimates presented there.

In Section 2, evidence is presented to show that the age of innovators at the time of their great inventions is increasing. Several possible reasons for this trend are discussed, including demographic shifts in the age distribution of the underlying population. Based on this discussion, Section 3 develops a formal stochastic model of the expected age at which inventors make their great inventions. This model accounts for (1) shifts in the age distribution of the population, (2) innovation potential over an inventor's lifecycle, and (3) the possibility that reaching the knowledge frontier has become more difficult over time. Parametric structure is imposed only on element (3), the thesis of special interest. The model assumes in particular that the age at which innovators arrive at the knowledge frontier is normally distributed and allows for a polynomial time trend in the mean of this distribution. The model is then estimated using semi-parametric maximum likelihood methods, with the results presented in Section 4. The results show a statistically significant upward trend in the estimated age at which innovators reach the knowledge frontier. The point estimate for the linear trend is 6.6 years per century, which is very similar to the age trend found in Chapter 1. On average, great inventors achieved the knowledge frontier at age 21.6 at the beginning of the 20th Century, but only at age 28.2 at the end. Section 5 discusses the implications of this result. Section 6 concludes.

2.2 Age Trends among Innovators

This section presents benchmark age trends among four groups of notable 20th-century "innovators": Nobel Prize winners in physics, great inventors, Pulitzer Prize winners in poetry, and Pulitzer Prize winners in drama. The former two groups are, of course, particularly relevant to economic growth, while the latter two groups are included for comparison. The age trends are determined by regressing, on the innovation year, the age of the innovator at that time. The innovation year was determined using biographical sources. Depending on the group and individual, "innovation" may mean date of crucial research, journal article, patent, or copyright. The Data Appendix describes the data sources. Table 1 presents the age trends and other summary information.

Table 2.1: Age Trends among Various Groups of Inventors

	Nobel Prize Winners in Physics	Great Inventors	Great Inventors (American)[a]	Pulitzer Prize Winners in Poetry	Pulitzer Prize Winners in Drama
Age trend[b]	5.97** (2.61)	4.94** (2.31)	7.26** (3.48)	24.5*** (5.74)	4.07 (5.14)
Number of observations	132	286	140	66	69
Time span	1881-1986	1900-91	1901-88	1917-99	1917-97
Average age	36.2	39.0	38.4	46.9	41.5
R^2	0.0386	0.0158	0.0307	0.222	0.0093

[a] This column presents the American subset of the Great Inventors data presented in the prior column.
[b] Age trends are measured in years per century. Standard errors are given in parentheses.
** Indicates significance at a 95% confidence level.
*** Indicates significance at a 99% confidence level.

The first result of note in Table 1 is that all these groups are showing upward age trends. With the exception of Pulitzer Prize winners in drama, these upward age trends are significant. Note as well that the trend among Nobel Prize winners in physics and great inventors is remarkably similar – about five or six years over a century. The agreement between these two independent groups suggests not only a greater robustness of the age trend, but also the possibility of a difference-in-difference style analysis. If we view the scientific/technological innovators as a "treatment" group that may be experiencing the effects of knowledge accumulation, then we might profitably attempt a comparison with "control" groups that are claimed on a priori grounds to be immune from knowledge accumulation. Innovators in the creative arts, whose creations might be thought to be less dependent on earlier work, could form such control groups. As can be seen in Table 1 however, Pulitzer Prize winning poets show an enormous upward age trend over the history of that prize – some 25 years over a century. While this result could be interpreted within the difference-in-difference framework, the most reasonable conclusion may be simply that the a priori assumption was incorrect. Certainly, the treatment and control groups likely have many differences outside of the influence of knowledge accumulation. The comparison is included in this chapter to help motivate a deeper search for the underlying forces that drive the age trends.

One possible ingredient in any age trend among innovators is shifts in the underlying age distribution of the innovators themselves. If the population of innovators is getting older on average, then, ceteris paribus, the average age of innovators will be rising. A second ingredient is possible shifts in the "innovation potential" of individual innovators as a function of their age. Innovators' effort levels at a given age may be rising or declining over time as a function of many possible economic, cultural, medical or other factors. For example, medical advances may improve innovators' ability to continue their work at later ages; on the other hand, rising wealth or other factors may lead to less effort at later ages – we know for instance that retirement ages in the United States and other leading economies are declining (Costa, 1998). The issue of knowledge accumulation enters through this concept of innovation potential. If knowledge accumulation is creating an increasing training burden on innovators, we would expect to see reduced innovation potential at younger ages.

The following section will place these ingredients into a formal model which will then be estimated in Section 4. But first, to close our initial discussion of innovators' age trends, we should consider two further issues. The upward age trend among Pulitzer Prize winning poets is so large that it might seem difficult to explain on the basis of our discussion heretofore. One additional source of the trend could be selection bias: while the Pulitzer Prize, like the Nobel Prize, is given for specific accomplishments, selection committees may have begun considering lifetime achievement more as the award has matured and become more prestigious. A trend towards lifetime achievement could create an upward trend in age. Whether such a bias exists or is important is difficult to assess. The possibility of its existence, however, provides one reason to focus on the great inventors data set in our eventual empirical estimations. That data set, which is drawn from technological almanacs, appears more immune to possible selection issues.

The editorial boards' purposes were to select technological advances as opposed to individuals; it is harder to imagine a compelling story for how the selection of innovations would create bias in the innovators' ages.

A further issue relates to the theories of knowledge accumulation we are trying to identify. In the long-run, the crucial question is whether knowledge continues to accumulate or whether there are, periodically, revolutions in scientific thought that relieve innovators from the knowledge burden of past research. Whether knowledge is fundamentally cumulatory or revolutionary in nature is ultimately what we would like to test. In the short run, however, the Kuhnian model of revolutionary knowledge is still potentially consistent with an upward age trend – we could simply be witnessing a period of within-paradigm innovation as opposed to a period of paradigm shifts. Such a view might explain the upward age trend among physicists. However, such a view is harder to maintain across a multidisciplinary data set. The great inventors data set thus has a second advantage over the others: it is constituted by noted innovations across many fields. A final reason to focus on the great inventors is that this group is most relevant to economic growth.

2.3 Econometric Model

The arrival of new ideas to innovators is likely best modeled as a stochastic process. Therefore, the age at which great innovators produce their great inventions can be viewed as a random draw from a distribution. The empirical goal is then to identify that distribution and, in particular, to identify how knowledge accumulation can influence that distribution. Two ingredients to the distribution are clear: (a) the age distribution of the population from which the innovators are drawn; (b) the innovation potential of innovators as a function of their age. Different models of knowledge accumulation can be embedded in this stochastic model through their influence on element (b), particularly by influencing the innovation potential of innovators at younger ages. This section presents a simple stochastic model to define the probability that witnessed innovations are produced by innovators at particular ages. The empirical analysis will use this model to generate a maximum likelihood estimate of any trend in the mean arrival age at the knowledge frontier.

Formally, consider a population N. Given that we have witnessed an innovation, the probability that the innovation was produced by an individual i is defined by:

$$\Pr(i) = \frac{x_i}{\sum_{\{i \in N\}} x_i}$$

where x_i represents the innovation potential of person i. Innovation potential can be interpreted as the instantaneous probability that person i produces an innovation. However, innovation potential need not be given such a stochastic basis; it need only be understood as measuring the relative innovative strength of an individual.[3]

It will be useful to consider the model in terms of cohorts of equally-aged individuals. First, define the set of cohorts as A, where $a \subset A$ represents the cohort with age a. Furthermore, let the set of individuals in this cohort be $N_a \subset N$, and let the number of individuals in such a set be defined as $|N_a|$. Then the average innovation potential of individuals in the cohort with age a is,

$$\bar{x}_a = \frac{1}{|N_a|} \sum_{\{i \in N_a\}} x_i \tag{2.1}$$

By immediate extension, the probability that a witnessed innovation is produced by an individual in the cohort with age a is,

$$\Pr(a) = \frac{\sum_{\{i \in N_a\}} x_i}{\sum_{\{i \in N\}} x_i} = \frac{|N_a|\,\bar{x}_a}{\sum_{\{a \in A\}} |N_a|\,\bar{x}_a}$$

Dividing top and bottom by the size of the entire population, $|N|$, and defining the age distribution of the population as $p_a = |N_a| \,/\, |N|$, we can again rewrite this expression into a particularly

---

[3] One may think of innovators as being drawn, with replacement, from a box of names. A particular person's innovation potential then represents the frequency with which his or her name appears in the box, where we imagine that innovators with higher ability or effort level appear more often.

useful form,

$$\Pr(a) = \frac{p_a \overline{x}_a}{\sum_{\{a \in A\}} p_a \overline{x}_a}$$

To this point we have ignored the possibility in our notation that both the age distribution of the population, p_a, and the average innovation potential of individuals of age a, $\overline{x}_a$, may be changing over time. Investigating this possibility is of course the ultimate purpose of our estimation exercise and we acknowledge it explicitly be writing,

$$\Pr(a|t) = \frac{p_a(t)\overline{x}_a(t)}{\sum_{\{a \in A\}} p_a(t)\overline{x}_a(t)} \tag{2.2}$$

Any variation in the expected age of innovators over time is then determined explicitly:

$$E[a|t] = \sum_{\{a \in A\}} a \Pr(a|t)$$

More generally, equation (2.2) provides the central vehicle for the maximum likelihood estimation presented in Section 4.

Several important if obvious points are worth making explicitly. First and foremost, it is now clear that any (non-random) trend in the age of innovators must be driven either by trends in the population age distribution, $p_a(t)$, or by trends in the average innovation potential of various age groups, $\overline{x}_a(t)$. Second, any presumption that the innovators' upward age trends are driven by increasing life expectancy may be misleading. Increases in life expectancy are unlikely to be an important force, simply because the innovation potential of those in their later years is likely to be low – if only because people retire. Furthermore, demographic events like the post-war baby boom in the U.S. can create periods where the working population is getting younger even though life expectancy is rising. Third, the stochastic process represented in equation (2.2) can produce innovators with a large variance in age. This fact explains why we see low values for R^2 in the trends presented in Table 1. Finally, it is worth noting that the stochastic model derived above makes very few assumptions.[4] While there will be plenty of room to question how $p_a(t)$ and $\bar{x}_a(t)$ are estimated, the model to this point is quite general.

The next step in the model derivation, and where we will begin to make more stringent assumptions, regards the definition of x_i, and, by extension, $\bar{x}_a$. In particular, we need to embed in equation (2.2) a sub-model that allows for trends in the age at which innovators reach the knowledge frontier. A simple approach is to assume that innovators start their lives with a period of education during which they have zero innovation potential. Let the (stochastic) length of education required for an individual i be e_i.[5] Additionally, we will define $g(a_i; z_i)$ as the individual's innovation potential if fully educated, where z_i is some (stochastic) measure of talent, effort, health, and any other factor that influences innovation ability. The innovation potential of individual i as a function of their age is then,

$$x_i = I(a_i \geq e_i) g(a_i; z_i)$$

where $I(a_i \geq e_i)$ is an indicator function equal to 1 if $a_i \geq e_i$ and 0 otherwise.

With this specific description of individual innovation potential, we can employ a law of large numbers to write the cohort average innovation potential as,

$$\bar{x}_a(t) \xrightarrow{p} E[I(a_i \geq e_i) g(a_i; z_i)]$$

Making the additional assumption that that e_i and z_i are independent[6] this expectation simplifies to,

---

[4] Two assumptions the derivation does implicitly make are that (i) cohort innovation potential is independent of the nature of the innovation and (ii) there are no spillovers between innovators

[5] Two reasons we might expect a distribution in the age at which innovators arrive at the knowledge frontier are (i) different fields contain different types or amounts of knowledge, and (ii) individuals vary in the speed at which they educate themselves within a specific field.

[6] In general, it might be more reasonable to assume that the distributions of e_i and z_i are not independent: high ability (high z_i) individuals are likely to be more efficient in their education (low e_i). While this possibility is not considered formally in this paper, the data set used here considers great inventors, for whom we might take the view that z_i is essentially constant.

$$\bar{x}_a(t) \xrightarrow{p} \Pr(a_i \geq e_i) E[g(a_i; z_i)] \tag{2.3}$$

We will estimate the expectation of $g(a_i; z_i)$ non-parametrically using data on the life-cycle output of great innovators (see Section 4.2). To estimate $\Pr(a_i \geq e_i)$, the probability that an innovator of age a_i in a particular cohort has reached the knowledge frontier, we will impose some parametric structure. We will assume first that e_i is distributed normally within cohorts,

$$e_i \sim N[\mu(t), \sigma(t)] \tag{2.4}$$

The key question of interest is whether $\mu(t)$, the mean age of arrival at the knowledge frontier, is trending over time. Shifts in this mean over time can be generally modeled by a polynomial expansion,

$$\mu(t) = \theta_0 + \theta_1 t + \theta_2 t^2 + ... \tag{2.5}$$

Shifts in the standard deviation of the normal c.d.f. can be modeled similarly. The estimations presented in Section 4 assume that the standard deviation is fixed over time and limits equation (2.5) to linear and quadratic forms. Taken together, equations (2.2), (2.3), and the distributional assumptions in (2.4) and (2.5) produce a stochastic model that integrates demographic effects with a model of knowledge accumulation. We can now attempt to determine what is driving the upward age trend among inventors.

2.4 Estimation and Results

This section proceeds in several parts. It first discusses the data and procedures used to estimate $p_a(t)$ and $\bar{x}_a(t)$. It then turns to estimates of the sources of the inventors' age trend, presenting this chapter's central results, and concludes by considering the robustness of the estimations.

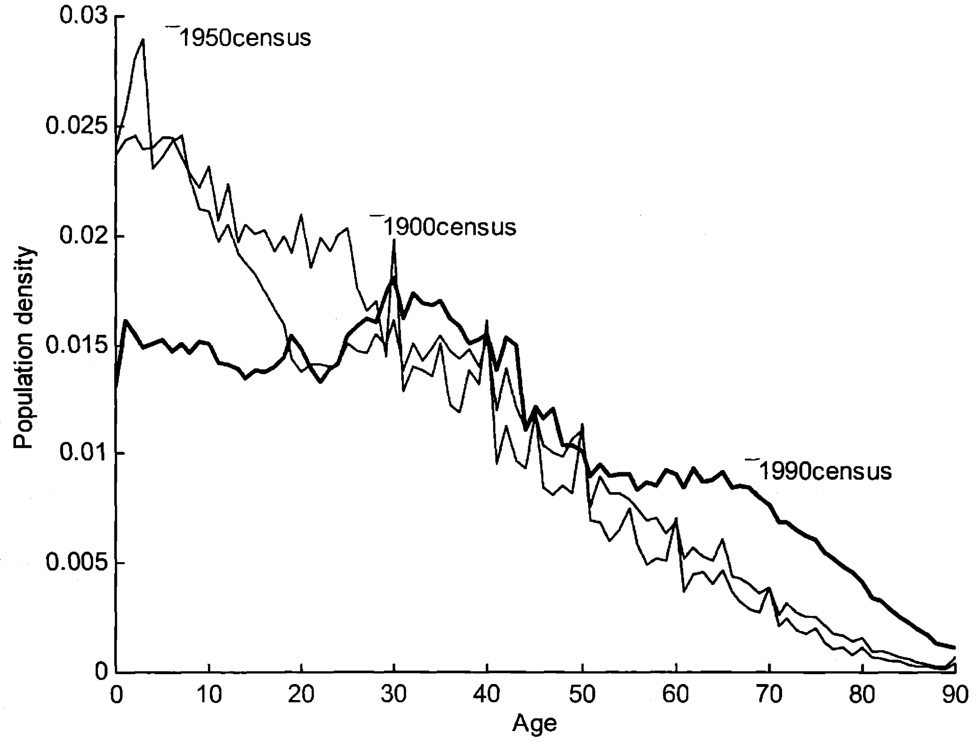


Figure 2-1: Age Distribution of U.S. Population for Various Census Years

2.4.1 Choosing populations and estimating age density

The inventors in the great inventions data set come from many different countries and are therefore drawn from populations with differing age distributions. Data on these age distributions are difficult to find for many countries, particularly over the timeframe of the entire 20th century. For this reason, the estimation will focus on the American subset of great inventors, who show a similar trend in age as the larger group and who provide a significant number of observations on their own. See column (3) in Table 1.

Figure 1 shows the population age density, $p_a(t)$, of the United States for three selected census years. The densities are calculated from large micro-samples of the U.S. census.[7] With

---

[7] The micro-samples are available electronically through the University of Minnesota. The census data are available on a decennial basis. Data for the years in between, which will be needed to estimate the model, are

these micro-samples, it is possible to determine both the age distribution of the national population as well as age densities for various subgroups, such as "workers" or subgroups defined by specific occupational types. Note that the model developed in Section 3 can be interpreted to use either type of data. If we define the population (N) and hence $p_a(t)$ to include the entire U.S. population, then we are implicitly including in each cohort of "innovators" a large number of people with zero or nearly zero innovation potential. If the percentage of the U.S. population at a given age who are active innovators is trending over time (for example, because students in successive cohorts are becoming more or less interested in innovative careers) then the age distribution of the national population does not accurately represent the age distribution of the sub-population of innovators. We could avoid this potential pitfall to some degree if we were able to define the population to include only those individuals who are "true" innovators. However, while the census data do allow one to break out specific occupational categories, they cannot help here for two reasons. First, the U.S. Census's definitions of occupational categories are neither sufficiently specific nor consistent over the 20th century to identify the subgroup of innovators with any confidence. Second, occupations are not defined until one begins them; therefore, using occupational based age distributions from the U.S. Census would cause evidence on the effects of knowledge accumulation to evaporate into $p_t(a)$.[8] For the estimation exercise in this chapter, the entire U.S. population will be used to estimate $p_t(a)$. This assumption will receive further consideration later, and we will see that defining the age distribution of innovators this way is likely conservative.

2.4.2 Estimating innovation potential

The next task is to produce a non-parametric estimate for $E[g(a_i, z_i)]$. To do this, data were collected on 31 long-lived individuals, all of whom were born in the mid to late 19th century.

---

generated by linear interpolation. The census data are further discussed in the Data Appendix.

[8] This raises the possibility of using the Census data directly to estimate trends in the age of arrival at the knowledge frontier; however, given the bluntness and inconsistency of the Census occupational categories with regard to innovative careers, such an exercise is unlikely to produce convincing interpretations; it also loses the focus on great inventors.

Eleven of the individuals are inventors who hold significant numbers of patents. Their innovation potential, $g(a_i, z_i)$, was defined by patent frequency as a function of age. The remaining individuals are eminent American scientists across a variety of disciplines. Their innovation potential was defined by publication frequency. The average innovation potential is presented in the figure along with a smoothed kernel estimate that uses a bandwidth of six years.[9]

Figure 2 provides an estimate of average lifetime innovation potential, $\bar{x}_a$. Two salient features in the figure are: (1) a significant period of zero innovative output during childhood followed by a steep climb in innovation to the age of forty; and (2) a more gradual tapering in creative output beyond the average innovator's mid-fifties. It is essential to note that the averages presented in Figure 2 are conditional on being alive, so that the tapering later in life is not due to a dwindling sample size but rather to a decrease in effort and/or ability. Note also that the kernel estimate for $\bar{x}_a$is very similar in shape to that found in the psychology literature.[10]

The estimate of $\bar{x}_a$ in Figure 2 combines both its subsidiary elements, $\Pr(a_i \geq e_i)$ and $E[g(a_i; z_i)]$ (see equation (2.3)). The initial part of the curve in Figure 2 clearly indicates that innovators at young ages, $a_i < e_i$, are not actively producing innovations. This feature of the graph obscures what their raw innovative potential, $g(a_i; z_i)$, might be were they already educated and actively producing innovations. For estimation purposes, it is necessary to assume some shape of $E[g(a_i; z_i)]$ at these young ages where it is impossible to witness it directly. A reasonable way to proceed is to assume that $E[g(a_i; z_i)]$ at ages below the peak of the kernel estimate is equivalent to that at the peak. This says that innovators' raw capacity to innovate at young ages would be equivalent to their highest witnessed performance were the young

---

[9] See the Data Appendix for further details.

[10] Simonton provides a cognitive model of life-cycle creative productivity which takes the form:

$$x_i = a\left(e^{-bt} - e^{-ct}\right)$$

where t measures career time as opposed to age. With parameter values he suggests, the structural features of this model are similar to what we see in Figure 2. Namely, innovative potential rises steeply in one's 20's, peaks in the late 30's or early 40's, and then trails off more slowly in later years. These features have been shown to fit a variety of empirical data extremely well (Simonton, 1991).

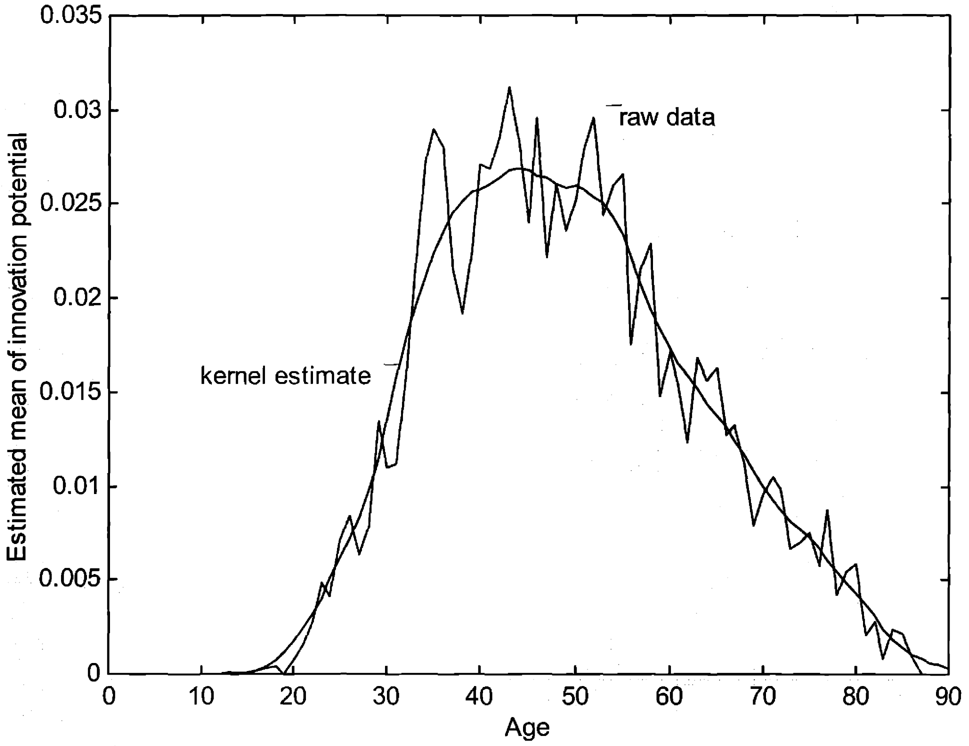


Figure 2-2: Estimate of Life-cycle Innovation Potential

innovator not constrained by educational requirements.

The maximum likelihood estimations presented below further assume that movements in innovation potential over time only occur on the rising part of the innovation potential curve. That is, the estimations assume that $E[g(a_i, z_i)]$ does not change over time, so that any shift in $\bar{x}_a(t)$ occurs because the distribution of e_i, the educational requirement, is changing. While this is a significant simplification, we will argue later that it is probably a conservative one.

2.4.3 Central results

Panel 1 of Table 2 examines whether population shifts alone might explain the age trend among American inventors. In all three specifications, innovation potential is assumed to be fixed over time. The expected age of the innovator is calculated using equation (10) and presented for several different years, together with the expected age trend over the course of the 20th century.

The first specification presumes that innovation potential is constant over the lifecycle. With this assumption, equation (10) simply calculates the average age of the U.S. population. We see that the average age of an American has risen monotonically by nearly 10 years – a trend larger than the age trends we have found among innovators. Increasing life expectancy is the obvious principal force. The second specification presumes that innovation potential is constant between the ages of twenty and sixty-five, and zero otherwise. Equation (10) therefore calculates the average age of the American workforce (somewhat crudely defined). The final specification uses the kernel estimate of innovation potential presented in Figure 2. With this specification, equation (10) calculates this chapter's best guess at the expected age of an innovator. Two important features emerge. First, the expected age trend is only four years now, smaller then the age trends we witnessed among inventors. Second, the expected age trend is no longer monotonic but rather flattens out between 1950 and 1990, which is due to the effect of the post-war baby boom. In sum, shifts in population age density do not appear to fully explain the upward age trends among innovators.

Panel 2 of Table 2 presents the semi-parametric maximum likelihood estimates of the knowledge accumulation model. The likelihood function is estimated assuming that each inventor is an independent draw from the distribution in equation (9). Innovation potential and population density are estimated as discussed above.

The first specification presents the central estimates of this chapter. Constraining the knowledge accumulation model presented in Section 3 to have a time-invariant standard deviation and a linear trend in the mean, we find that the average age at which innovators arrived at the knowledge frontier was 21.6 years of age in 1900 and has risen by some 6.6 years over the course of the 20th century. The linear upward age trend is significant at a 95 percent confidence level. Maximum likelihood standard errors are calculated using the covariance matrix of the first derivatives vector.

Table 2.2: ML Estimates of Knowledge Accumulation.

Panel 1: Expected ages and trends for different specifications of x(a)

		Specifications[a]	
	(1)	(2)	(3)
$E_{1900}[a]$	25.5	36.7	42.3
$E_{1950}[a]$	27.5	38.7	43.5
$E_{1970}[a]$	31.8	40.4	46.2
$E_{1990}[a]$	35.1	39.7	45.2
Age trend[b]	9.60 (1.40)	3.83 (0.94)	4.18 (0.78)

[a] Specifications: (1) x(a) = constant
(2) x(a) = constant if 20 ? a ? 65, 0 otherwise
(3) x(a) = kernel estimate presented in Figure 2.

[b] Age trends are measured in years per century. Standard errors are given in parentheses.

Panel 2: Maximum likelihood estimates of knowledge accumulation model

	Specifications[c]	
	(1)	(2)
θ_0	21.6 (1.64)	21.7 (4.08)
θ_1	6.65 (3.13)	6.44 (20.1)
θ_2	--	0.00 (0.20)
σ	3.21 (1.16)	3.21 (1.18)

[c] Specifications: (1) Linear trend model
(2) Quadratic trend model

The second specification allows for a somewhat richer estimate of the trend in the mean by including an additional parameter to measure any quadratic element. We see that the standard errors on the trend parameters grow so large that we lose significance, implying that the data may be too few to support the richer parametric specification. However, the quadratic estimation does produce point estimates that are nearly identical to those in the linear case.

These maximum likelihood estimates are consistent with the possibility that the knowledge requirements faced by innovators are growing ever more burdensome. Taking the point estimate, inventors today require 6.6 years longer to reach the knowledge frontier than they did at the beginning of the 20th century. Remarkably, this estimate is nearly identical to the age trend estimate produced in Chapter 1 of this thesis, where the trend in age at first innovation was estimated directly for a large sample of innovators for the shorter period between 1975 and 2000. Such a trend implies, ceteris paribus, a monotonically decreasing return to society's investment in R&D. It also implies, ceteris paribus, a monotonically decreasing private return to the potential innovator. Section 5 will briefly highlight the main implications of such a trend and connect them back to growth theory. First, we will consider three possible objections to the identification strategy this chapter has employed.

2.4.4 Robustness

The amalgamation of so many important factors in the innovation potential function will leave this model open to the criticism that the effect of knowledge accumulation within it is not well identified. Effectively, we have set up a contest between population trends, represented in $p_a(t)$, and knowledge accumulation, represented through $\bar{x}_a(t)$, to determine which explains the upward age trends among inventors. As previously noted, we might be concerned that the U.S. population data used to estimate $p_a(t)$ do not adequately reflect the subgroup of potential innovators. Or we might be concerned that by employing a model that leaves $\bar{x}_a(t)$ effectively fixed in time for ages greater than forty, the model has prevented various effects that are prevalent in the latter half of the lifecycle from playing potentially important roles in explaining the age trends. Finally, we may be concerned that the upward trend found in the age that marks the beginning of a research career, even if the trend does exist, may have nothing to do with knowledge accumulation. This subsection will examine each of these concerns in turn.

With regard to population, existing evidence shows that the number of researchers in the United States has been growing since 1950 at a rate approximately four times that of employment as a whole (Jones, 2001). Assuming that the new innovators are mostly being produced as first careers and not by career switches or immigration, the faster growth in R&D employment will produce a body of innovators who are younger on average than the rest of the working population. However, if the higher growth rate of R&D employment has been consistent over a long period of time, then the distribution of the population of innovators will be stable. In other words, a constant growth rate in the number of innovators cannot alone produce a trend in the expected age of innovators. Concavity in the employment of researchers – that is, a slowdown in the growth of young R&D workers – is required to produce an upward age trend, but there is no evidence of a (non-cyclical) slowdown going back to 1950.[11] Data available from Machlup indicates that the R&D employment growth rate was constant between 1900 and 1954 (Machlup, 1962).

The second issue to consider is the assumption that $\bar{x}_a(t)$ is time invariant at later ages. In particular, one might be concerned that improvements in health have increased innovation potential disproportionately with age, allowing people to work longer and more effectively in their later years. However, retirement trends are pushing extremely strongly in the other direction. For example, while in 1900 some 91% of men aged 55-64 were working in the United States, only 84% were employed at those ages in 1960, and a mere 67% in 1990 (Costa, 1998). If retirement trends among innovators are at all similar to these trends for the population as a whole, it becomes difficult to imagine that cohort innovation potential at later ages has been rising over time. Keeping fixed $\bar{x}_a(t)$ at later ages appears to be a conservative assumption;

---

[11] U.S. R&D employment did experience an increased growth rate from the mid-fifties through the end of the 1960's, which was driven by increased defense expenditures and the space program (Jones, 2001). R&D employment then fell relative to the total labor force in the early 1970's.

it leads to an underestimate of the upward trend in the age at which innovators reach the knowledge frontier.

The final question concerns whether the upward trend in the implied age at first innovation says anything about knowledge accumulation at all. For example, cultural trends may be causing potential innovators to take more roundabout routes toward their eventual careers. Or perhaps educational institutions have become much less efficient with time. While it would seem difficult to support an argument that teaching quality has dropped substantially enough to explain the trend in age, an increased dependence on signaling in labor markets might be a more plausible alternative hypothesis. These types of arguments are difficult to assess within the context of this chapter. And it is important to note that evidence on the education of inventors would not help decide the issue. Increasing education is consistent with both an increasing need for knowledge accumulation, increasing incentives to signal quality through educational tenure, and decreasing educational efficiency. The identification strategy in this chapter implicitly assumes that innovators accumulate knowledge as a necessity and at a rate that is invariant with time. If the reader considers great inventors to be a group of highly talented and motivated individuals, then perhaps alternative stories based on signaling and/or declining educational efficiency seem less compelling.

2.5 Implications

The empirical trends uncovered in this chapter may have significant long-run implications. Rising educational demands imply, ceteris paribus, more time spent educating and less time spent innovating. This section will examine the potential impact of these effects and possible ways to avoid them.

One of the major insights of growth theory in the last decade rests on the idea that technological progress comes specifically from the employment of human resources. In both idea-based growth models and learning-by-doing models, growth rates will consequently increase with the number of people employed in the innovative sectors: the greater the number of heads, the greater the number of ideas. Recent empirical exercises have found support for the positive role of population in the rate of technological progress (e.g. Kremer 1993, Jones 1995).

The central implication of the research presented here and in the last chapter is that the research output of an economy will depend not just on the number of people employed in R&D but also on the individual innovative capacities of these individuals. A rising burden of knowledge will tend to reduce these innovative capacities. In Chapter 1, the abilities of individuals were seen to become increasingly limited due to narrowing expertise. In this chapter, the productivity of innovators is reduced because they spend a greater portion of their good years undertaking education as opposed to research. Either way, the implications for growth are similar – a given number of researcher-lives becomes less and less productive, and steady-state growth cannot be maintained. As discussed in Chapter 1, two basic ways to maintain steady-state growth despite an increasing knowledge burden are: (1) to continually put more and more people into R&D; and (2) to presume that the size or frequency of an individual's ideas is rising at or above the growth rate. The first solution will work indefinitely as long as populations can continue to grow. However, the available evidence suggests a negative relationship between fertility and income levels, so that populations may be unlikely to grow in the long run. The second solution is an exogenous feature of technological progress, outside policy control, and recent evidence (e.g. Jones 1995, Kortum 1997, Chapter 1 of this thesis) indicates that productivity contributions per innovator are falling. This paints a somewhat pessimistic picture.

The life-cycle considerations introduced in this chapter raise further, specific issues and areas for possible policy intervention. First, educational efficiency is clearly a key parameter. Efforts to improve the rate of learning – by enhancing the rate of information transmission to students and/or better packaging of existing information – will mitigate the effects of increasing educational requirements. This motive for educational efficiency adds further texture to existing debates over educational expenditures and institutional design, from teacher salaries to classroom size to the concept of a liberal arts education. Efficiency arguments may be further emphasized given the sentiments Dirac discussed in the introduction.

A second issue is lifespan and more generally health at later ages. Presumably, longer lives can lead to longer working lives and thereby compensate for longer periods of education at the beginning of the life cycle. Taking a long-run view, human lifespan is clearly increasing and is presumably at least mildly policy-sensitive, adjustable through subsidized research. Even so, increases in lifespan alone may be insufficient given morbidity concerns, so that cures for diseases that cause declining mental function in old age would be important and perhaps necessary complements to longer lifespans. Finally, as cited earlier, retirement ages are declining despite longer lifespans and better health care, implying that incentive problems among older workers may be at least as critical as lifespan and health concerns in shaping the timespan of employment.

2.6 Conclusions

This chapter considers further empirical evidence for a rising burden of knowledge. The chapter begins by establishing that great inventors are getting older at the time of their great inventions. The estimates suggests that great inventions are produced at ages 5 or 6 years later at the end of the 20th Century than they were at the beginning. Noting that this trend may occur simply because the population is aging, the chapter derives a formal econometric model that incorporates the shifting age structure of the population together with the thesis of interest – that increasing learning requirements are delaying the innovative portion of researcher's careers. Semi-parametric maximum likelihood estimation is then employed to ascertain the source of the upward age trend. The model estimates that, while the U.S. population has been getting older, the average age at which great inventors have been reaching the knowledge frontier has also been increasing. The model estimates a linear increase of 6.6 years in the age at which innovators achieve the knowledge frontier. This estimate is nearly identical to the trend found in Chapter 1 for the age at first innovation. The research in the current chapter suggests that such age trends describe not just the recent past but the 20th Century as a whole. Furthermore, the new results suggest that a rising educational burden is not just a constraint for the average patent holder; it also falls on the shoulders of the greatest minds.

The growth implications of a rising burden of knowledge may be significant. Greater educational requirements reduce the lifetime capacity of individual innovators. If knowledge accumulates deterministically as technology advances, then pessimistic growth implications may be unavoidable. Policies for improving human capital – whether improving education for the young or health for the old – could mitigate a rising knowledge burden.

The research in this chapter (and the last) is also relevant to debates over the nature of scientific progress. Thomas Kuhn, the historian of science who coined the word "paradigm" in its modern usage, began arguing in the 1960s that scientific progress is marked in the long run by fundamental revolutions – "paradigm shifts" – as opposed to a smooth, continuous accumulation of scientific truth (Kuhn, 1962). Scientists, disinclined to believe that current views are fundamentally wrong, tend to prefer the smooth accumulation view, which might seem to be a more optimistic description of human progress. Here an interesting irony appears: when we look at the issue through the lens of economic growth, the revolutionary view appears far more optimistic.

Further historical research along the lines seen in this chapter can attempt to better incorporate or attenuate the robustness concerns outlined in Section 4.4. Additional data on the employment of innovators would help address concerns that measures of population age density used in this chapter's estimations do not adequately represent the age distribution of the population relevant to the innovative process. Other data might enrich our understanding of the forces that influence innovation potential later in the lifecycle. Finally, data on great inventions and their inventors could be expanded both cross-sectionally and still further back in time.

2.7 Data Appendix

This appendix describes the data sources used in this chapter, providing both reference material and some underlying details of the methodology used in data collection.

2.7.1 Data on innovators

There is a wealth of biographical information available on the Nobel Prize winners in physics, and several sources were used to obtain dates of birth, the reason for the prize, and the year in which the prize-winning research was performed. The last piece of information, which can be the most difficult to ascertain, was taken either to be the appearance of the key journal article or the final year of evident work on the key research. Cross-referencing the sources listed below, I was able to rate a subjective confidence level of the year in which the prize-winner made their specific contribution to the field. Regressions on subsets defined by confidence level produce virtually the same trend as for the group as a whole; the data and trend reported in Table 1 are for the entire group. The primary source was,

Schlessinger, B. and Schlessinger, J. *The Who's Who of Nobel Prize Winners*, 1901-1995. Oryx Press, Phoenix AZ 1996.

which was cross-referenced with,

Daintith, J. and Gjertsen, D. *The Grolier Library of Science Biographies.* Vols. 1-10. Grolier Educational, Danbury CT 1996.

Debus, A.G. ed. *World Who's Who in Science: A Biographical Dictionary of Notable Scientists from Antiquity to the Present.* Marquis Who's Who Inc., Chicago 1968.

McMurray, E.J., Kosek, J.K., and Valade, R.M. *Notable Twentieth-Century Scientists.* Vols. 1-4. Gale Research, Detroit 1995.

Williams, T.I. ed. *Biographical Dictionary of Scientists.* John Wiley and Sons, New York 1974.

Data on great inventors were collected from two technological almanacs that provide, by year, a list of notable technological advances. These almanacs typically provide the date and location of birth of the innovator responsible. The almanacs used were,

Bunch, B. and Hellemans, A. *The Timetables of Technology.* Simon and Schuster, New York 1993.

Ochoa, G. and Corey, M. *The Timeline Book of Science.* Ballantine Books, New York 1995.

In the event that the almanacs did not provide dates or location of birth, the four general biographical resources listed above were consulted, in addition to two further biographical resources:

Bachman, T.M. ed. *Who's Who of Science and Engineering: Millennial Edition.* Marquis Who's Who, Wilmette, IL 1999.

Inventors' Hall of Fame. Website: http://www.invent.org/. Akron, OH.

Data on the Pulitzer Prize winners in drama and poetry were collected from a single source that provided all the required information,

Brennan, E. A. and Clarage, E. C. *Who's who of Pulitzer Prize winners.* Oryx Press, Phoenix AZ 1999.

For the poets, the year of innovation was defined as the year of copyright for the award-winning collection of poetry. For the dramatists, the year of innovation was defined as either the year of copyright or the year of the first performance, whichever came first.

2.7.2 Data on age distribution

One and five percent micro-samples of the U.S. census are available electronically through IPUMS, the Integrated Public Use Microdata Series, which is maintained by the University of Minnesota. The smallest sample used was for the 1900 census, whose micro-sample provided data on approximately 100,000 individuals. The largest sample used was for the 1990 census, whose micro-sample provided data on approximately 2.5 million individuals. Existing census research available on the website (www.ipums.umn.edu/usa/chapter3/chapter3.html) indicates that these micro-samples provide accurate estimates of the population at large with regard to age.

2.7.3 Data on innovation potential

Patent data for 9 of the 11 patentees used in estimating lifecycle innovation potential were drawn from the following source,

Kraeuter, D.W. *Radio and Television Pioneers: A Patent Bibliography.* Scarecrow Press, Metuchen NJ 1992.

Patent data for the remaining 2 patentees, as well as data on the publication frequencies for 20 eminent American scientists, were drawn from multiple volumes of the following series,

National Academy of Sciences, *Biographical memoirs*, Washington, D.C.

The year in which a patent was granted was used to determine the age of the patentee at the time of innovation. The year of publication was used for the scientists. For the scientists, publications of any kind were included and equally weighted, whether they be journal articles, books, or lighter articles or editorials.

All 31 persons used in the (Epanichnikov) kernel estimation of innovation potential lived into old age. The two youngest at their death were 67. Seventy percent lived beyond the age of 79, most well into their 80's Given the differing ages of death, a simple average of innovation potential across these individuals would produce a misleading result. The averaging method employed in producing Figure 2 scales up the innovation potential of those who died at younger ages by presuming they would have produced a fraction of additional innovations beyond their death equivalent to the fraction produced by those individuals who outlived them.

Kernel estimations of the patentees alone or of the scientists alone produce estimates that are less smooth but very similar to each other in their larger features.

Bibliography

[1] Costa, Dora L. *The Evolution of Retirement: An American Economic History, 1880-1990*, Chicago: University of Chicago Press, 1998.

[2] David, Paul, "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," *American Economic Review*, May 1990, 80 (2), 355-61.

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

[4] ——. "R&D-Based Models of Economic Growth," Journal of Political Economy, August 1995, 103, 759-84.

[5] ——. "Sources of U.S. Economic Growth in a World of Ideas," July 2000. Working paper.

[6] Kortum, Samuel S. "Research, Patenting, and Technological Change," *Econometrica*, November 1997, 65, 1389-1419.

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

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

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

[10] Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, October 1990, 98 (5), S71-S102.

[11] Rosenberg, Nathan and Frischtak, Claudio R., "Long Waves and Economic Growth: A Critical Appraisal", *American Economic Review*, May 1983, 73 (2), 146-51.

[12] Schumpeter, Joseph A., *Business Cycles, a Theoretical, Historical and Statistical Analysis of the Capitalist Process*, New York: McGraw-Hill, 1939.

[13] Simonton, Dean K. *Scientific Genius: A Psychology of Science*, Cambridge: Cambridge University Press, 1988.

[14] ——. "Career Landmarks in Science: Individual Differences and Interdisciplinary Contrasts," *Developmental Psychology*, January 1991, 27, 119-130.

[15] ——. "Age and Outstanding Achievement: What Do We Know After a Century of Research?" *Psychological Bulletin*, September 1988, 104, 251-267.

[16] Solow, Robert M., "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, February 1956, 70, 55-94.

[17] Young, Alwyn, "Growth without Scale Effects," *Journal of Political Economy*, 1998, 106 (1), 41-63.

Chapter 3

Do Leaders Matter? National Leadership and Growth in the Developing World[1]

> "The historians, from an old habit of acknowledging divine intervention in human affairs, look for the cause of events in the expression of the will of someone endowed with power, but that supposition is not confirmed either by reason or by experience."
>
> – Leo Tolstoy
>
> "There is no number two, three, or four... There is only a number one: that's me and I do not share my decisions."
>
> – Felix Houphouet-Boigny, President of Cote D'Ivoire (1960-1993)

3.1 Introduction

In the large literature on cross-country economic performance, economists have given little attention to the role of national leadership. While the idea of leadership as a causative force is as old if not older than many other ideas, it is deterministic country characteristics and relatively persistent policy variables that have been the focus of most econometric work.[2]

---

[1] **This research has been performed jointly with Ben Olken.**

[2] See, for example, Sachs & Warner (1997) on geography, Easterly & Levine (1997) on ethnic fragmentation, La Porta et al (1999) on legal origin, and Acemoglu et al (2001) on political institutions.

A smaller strand of the literature has recently suggested a more volatile view of growth. The correlation in growth rates across decades ranges between only 0.1 and 0.3 within countries (Easterly et al, 1993). This weak correlation suggests that countries are, at different times, in substantially different growth regimes, and recent econometric work has helped to further substantiate this view (Pritchett, 2001; Jerzmanowski, 2002). Particularly for developing countries, growth is neither consistently good nor consistently bad. Rather, developing countries tend to undergo substantially different growth episodes that can last for years or decades.

To take an important example, consider post-war growth in China. Figure 1 plots the log of real PPP gross domestic product over time. It is quite clear from the graph that China moved from a low-growth regime to a high-growth regime in or around 1978. Growth between 1950 and 1978 averaged 1.6% per year, while growth since 1978 has averaged 6.3%. To understand the development experience of China, one wants to know what caused this dramatic shift. The answer is not likely to be found – for China or the many other countries that exhibit such shifts – in the slow-moving explanatory variables typically used in the cross-country growth literature. Shocks and/or high frequency events can presumably provide more obvious explanations. The purpose of this chapter is to examine the role of one possible force that changes sharply and at high frequency: the national leader.

Even casual observers of Chinese history might immediately notice a coincidence between the low-growth period in China and the rule of Mao Tse-Tung. Mao came to power in 1949 and remained the national leader until his death on September 9, 1976. The forced collectivization of agriculture and later, in the mid-1960's, the Cultural Revolution were among many national policies that likely served to retard growth during Mao's tenure. Arguably, Mao himself – the individual – could be seen as a powerful causative force. This type of interpretation is often described as the Great Man view of history, where events are best understood through the lives and actions of extraordinary individuals.[3] The antithesis, perhaps most prominently

---

[3] For example, the British historian John Keegan has written that the political history of the 20th Century can be found in the biographies of six men: Lenin, Stalin, Hitler, Mao, Roosevelt, and Churchill (Keegan, 2003).

associated with Leo Tolstoy, suggests that leaders are almost entirely subjugated to the various forces operating around them (Tolstoy, 1869). A more modern view in political science can point to the median voter theorem to suggest that national policy is not chosen by individual leaders. A modern view of leadership in the psychology literature considers the very idea of powerful leaders as a social myth, embraced to satisfy individuals' psychological needs (Gemmill & Oakley, 1999).

This chapter investigates whether national leaders can have a causative impact on national economic performance. Growth, the main object of explanation in this chapter, was chosen partly because of its general import and partly because it sets the bar for leaders very high. One might believe that leaders can influence various policies and outcomes long before one is willing to believe that leaders could impact something as significant as national economic growth.

To examine whether leaders can affect growth, one can investigate whether changes in national leaders are systematically associated with changes in growth. The difficulty, however, is that leadership transitions are often non-random, and in fact, may be driven by underlying economic conditions. For example, there is evidence in the United States that incumbents are much more likely to be reelected during economic booms than during recessions. (See, for example, Fair 1978 and Wolfers 2001.) Examining the impact of leaders on growth therefore requires identifying leader transitions that are unrelated to economic conditions or any other unobserved factor that may influence subsequent economic performance.

To solve this problem, we can again look to Mao as our guide. For a number of leaders, the leader's rule ended at death due to either natural causes or an accident. In these cases, the timing of the transfer from one leader to the next was essentially random, determined by the death of the leader rather than underlying economic conditions. These deaths therefore provide a natural experiment that can be used to examine whether leaders have a causative impact on growth.

This chapter uses a unique data set on leaders collected by the authors to examine the impact of leadership on growth. Data was collected to identify every national leader for 92 different developing countries in the post World War II period, from 1945 to 2000. For each leader, we also identified the circumstances under which the leader came to and went from power. For 53 of the leaders in the data, the leaders' rule ended by death due to natural causes or an accident. Using these 53 "random" leader transitions, we find substantial, robust evidence that leaders matter. Growth patterns change in a sustained fashion across these randomly-timed leadership transitions.

Given this result, an immediate question is whether there are common features of countries or political regimes that predict stronger leader effects. In particular, one might expect that leaders matter more in regimes with fewer restrictions on executive authority and with less developed institutions. We find evidence that the death of leaders in autocratic regimes leads to growth changes, while the death of leaders in more democratic regimes does not. We find further evidence that high settler mortality, which has been used as an instrument for lower levels of political institutional quality, also predicts where leaders are more likely to matter. We find no evidence, however, that a country's region, legal origin, or degree of ethnic fragmentation predict the degree to which leaders affect growth.

Finally, given that autocracies are the locus of strong leader effects, this chapter asks whether powerful leaders are good or bad for growth. We find that the deaths of autocrats lead to sustained improvements in growth, with point estimates suggesting a 2 to 3 percentage point improvement in growth rates over three, five, and seven year periods following the death.

The remainder of this chapter is organized as follows. Section 2 describes the leadership data set, and examines the 53 "random" leadership transitions in detail. Section 3 examines the impact of these leadership changes on growth, and presents the chapter's primary result: changes in leadership lead to changes in growth. Section 4 examines whether measures of institutional quality predict the degree to which leaders matter for growth. Section 5 considers the directional impact of leaders on growth and shows that the death of autocrats leads to substantial improvements in growth rates. Section 6 will consider a number of robustness checks on the results. Section 7 concludes.

3.2 Data

3.2.1 The leadership data set

This chapter uses a unique data set on national leadership collected by the authors. The focus is on the developing world, and the data set includes every nation in the Penn World Tables in Sub-Saharan Africa, South America, Central America, Asia, and the Middle East, for a total of 92 countries. For each country in the sample, we began with a list of all heads of state and heads of government in the 1945-1992 period, compiled from Lentz (1994). To extend this list of leaders through the end of the year 2000, we used data from the CIA World Factbook (2003) and the Zarate Political Collections (Zarate, 2003). The identity of each leader, their title, dates of tenure, and date of birth were assembled into a preliminary data set.[4]

The next step was to determine, at each point in the sample period for each country, which individual was the "national leader": the head of state, usually under the title of President, the head of government, usually under the title of Prime Minister, or perhaps some third figure. We defined the national leader to be the individual in the country who holds the most executive power, and determined the identity of this individual through extensive historical and biographical research.[5]

In most cases, identifying the national leader was straightforward, as most countries fell into one of four institutional structures with a clear national leader. In one set of countries, only one leadership position exists. This situation is particularly common in Latin America, where countries typically have presidents but no prime ministers. In the second set of countries, the same individual is both head of state and head of government. This situation is most common in dictatorial regimes and appears relatively often in Africa. In the third set of countries, the head

---

[4] For those countries that became independent after 1945, leadership data was only collected in the post-independence period.

[5] The major biographical sources used are listed at the end of paper.

of state is separate from the head of government, but one of the two is clearly subordinate to the other. Typically, the subordinate position is regularly appointed and dismissed by the other leader, and there are often interregnum periods in the subordinate role. This is particularly common in monarchies but holds in many other cases throughout the world. In the fourth set of countries, most often former British colonies, the head of state is a figurehead and power lies with the prime minister. Collectively, these four institutional settings, in which the national leader is clearly defined, account for 90% of the leaders in the sample.

Identifying the national leader in the remaining 10% of cases required further historical and biographical research. True institutional parity between the two roles is rare, so identifying which individual held the most executive power remained straightforward in most cases.[6] The resulting data set includes 685 different national leaders, representing 771 distinct leadership periods.

3.2.2 "Random" leader deaths

The leaders of particular interest for this chapter are those who died in office, either by natural causes or by accident. To define this group, further biographical research was undertaken to determine how each leader came and went from power. Table 1 presents summary statistics describing the departure of leaders. Of the 74 leaders who died in office, 21 were assassinated, 43 died due to natural causes, and 10 died in accidents.[7] We define the 53 leaders who died either of natural causes or in an accident as the "random" deaths that we focus on. Of these, heart disease is the most common cause of death, while cancer and air accidents were also relatively common. The most unusual death was probably that of Don Stephen Senanayake of Sri Lanka, who was thrown from a horse and died the following day from brain injury. Table 2 describes each of these 53 cases in further detail.

As will be discussed in more detail below, what is important for the identification strategy

---

[6] Military juntas, for example, often begin with a notionally rotating chairman, but such institutional arrangements do not last. An example of a more persistent, ambiguous situation is Thailand, where power over significant periods is held in a compromise arrangement between the military, the prime minister, and the king.

[7] Note that this excludes a further 18 leaders who were killed during a coup.

is that the timing of these 53 leader deaths be unrelated to underlying economic conditions. In particular, it is important that assassinations, which may be motivated by underlying changes in the country, be purged from the set of random leader deaths. For example, an important case is that of Presidents Habyarimana of Rwanda and Ntaryamira of Burundi, who were killed in the same plane crash in 1994. The ensuing Rwandan genocide produced a large shock to Rwanda's growth. Assassination theories are given some credence in this case and including Rwanda will – if we take the assassination view – bias the results toward finding a growth effect.[8] To reduce any such contamination of the data, each death episode was examined extensively using archived articles in major newspapers available through Lexis-Nexis. Conspiracy theories are for the most part uncommon, but Section 5, which considers various robustness checks on the results, will consider in detail the implications of conspiracy theories for the results. To be conservative, all analysis in this chapter will leave out the Presidents Habyarimana of Rwanda and Ntaryamira of Burundi.

3.3 Do Leaders Matter?

There are three reasons one might find an empirical relationship between particular leaders and particular growth episodes. First, leaders might impact their nations to the point that they can influence national growth. Identifying this possibility is the main purpose of this chapter. Second, in antithesis to the first, the causation might be reversed: changes in growth patterns may lead to leadership changes, perhaps because the electorate or powerful political groups demand change in response to economic events. Finally, an association between growth episodes and particular leaders may represent some third factor; for example, an increase in ethnic tension may produce shocks to both the political leadership and the economy.

Random leader deaths provide an opportunity to identify the causal impact of leaders on economic growth. Such deaths produce exogenously-timed shocks to the national leader, allow-

---

[8] The plane may have been shot down by artillery or a surface-to-air missile, though definitive evidence has never been produced publicly.

ing one to ask whether national leaders – as individuals – can impact the growth experience of their countries. This is not to say that a leader's power is independent of various institutional or other features that may condition the leader's impact. But randomly-timed leader deaths allow us to focus precisely on the role of the national leader and ask whether nations undergo substantial economic change when the leadership is changed. On average, randomly timed leadership changes will not be associated with other events that influence growth.

This section uses these randomly-timed leader transitions to show that leaders do, in fact, matter for growth. Section 3.1 provides a graphical overview of those countries with randomly-timed leader deaths. This analysis is informal but worthwhile; in many cases there are sharp, prolonged changes in national growth experiences when leaders die. Section 3.2 will present a formal econometric specification and consider the results of several growth regressions, comparing growth rates before and after leader deaths.

3.3.1 Graphical Evidence

Of the 53 identified episodes of random leader deaths, 8 occur in the period prior to the beginning of the Penn World Tables and therefore cannot be included in the growth analysis. As discussed above, we further exclude the deaths of Habyarimana of Rwanda and Ntaryamira of Burundi due to the likelihood that these deaths were assassinations. This leaves 43 random leader deaths, spread over 36 countries, for which there is comparable growth data. Figure 2 presents the log of real PPP gross domestic product over time for each of these countries. A solid vertical line represents the exact date at which a leader died. A dashed line represents the exact date at which that leader came to power. Cases where the entrance and/or exit from power occurs prior to the beginning of the Penn World Table observation period are not presented.

Looking at the graphs, it is clear that in a number of cases there is a sharp, prolonged change in the growth regime coincident with or just following the death of the national leader. This is particularly clear for Houphouet-Boigny in Cote d'Ivoire, Toure in Guinea, Khomeini in Iran, Machel in Mozambique, and, as already discussed, Mao Tse-Tung in China. Short-run changes in the growth pattern might also be seen in many other countries, including Angola, Egypt, India, and Nigeria, while subtler long-run changes might plausibly be seen surrounding leader deaths in several further cases, including Botswana, Gabon, Kenya, Pakistan, and Panama.

It is instructive to consider some of the more dramatic cases in further detail. The death of Machel led to an especially sharp turnaround in the economic performance of Mozambique (see Figure 2). Machel, the leader of the Frelimo guerrilla movement, became president in Mozambique in 1975 as Portuguese colonial rule collapsed. He established a one-party communist state, nationalized all land in the country, and declared free education and health care for all citizens. Coincident with Machel's aggressive policies, most Portuguese settlers fled Mozambique, and a new, debilitating guerilla insurgency was born. As is seen in Figure 2, Mozambique entered a sustained period of economic decline that continued throughout Machel's tenure. Upon Machel's death in 1986, his foreign minister, Joaquin Chissano, became the national leader. Chissano moved the country firmly toward free-market policies, sought peace with the insurgents, and established a multi-party democracy by 1990. Growth during Machel's tenure was persistently negative, averaging –7.7% per year; since Machel's death, growth in Mozambique has averaged 2.4% per year.

The case of Houphouet-Boigny of Cote d'Ivoire is more ambiguous. The sharp downturn in economic performance that began in the early 1980's is coincident with a collapse in the commodity prices for cocoa and coffee, Cote d'Ivoire's main exports. Houphouet-Boigny's death in 1993 shortly preceded a devaluation of the CFA, the regional currency shared by Cote d'Ivoire, which may have spurred growth by restoring the country's competitiveness in these products. However, one can also look to a number of policies associated with Houphouet-Boigny that appear poorly chosen: for example, his government borrowed and spent large sums in the 1980's to construct a new capital in Houphouet-Boigny's hometown of Yamoussoukro along with the world's largest catholic basilica, which would serve as his burial site.[9] In 1980, the

---

[9] This $300 million church was constructed from 1986-89, coincident with the arrest of striking government

Ivory-Coast had one of the highest per-capita incomes in Sub-Saharan Africa; in 1993, at the time of Houphouet-Boigny's death, it had experienced 14 consecutive years of economic decline, with growth rates averaging –3.0% per year.

The case of Ayatolloh Ruhollah Khomeini of Iran is more widely known. The Islamic Revolution in 1979 was followed by large-scale executions of opponents, international isolation over hostage-taking at the US Embassy, and a refusal to negotiate peace with Iraq despite massive losses of life and poor military prospects on both sides of the Iran-Iraq war. In particular, Khomeini cast the Iran-Iraq war in strict religious terms, which is said to have prevented any peace negotiations, although Iraq, having invaded unsuccessfully, withdrew from Iranian territory in 1982 and began seeking peace from that time. Iranian military tactics in the ensuing trench warfare included sending waves of unarmed conscripts, often young boys, against the superior firepower of entrenched Iraqi lines (Britannica, 2003). In the face of renewed Iraqi attacks, Iran finally accepted a UN brokered ceasefire in 1988, the year before Khomeini's death. Since his death, Iranian politics have become (relatively) more moderate; as can be seen in Figure 2, growth has turned substantially positive.

While these illustrations can provide some plausible interpretations in which leaders matter, such historical analysis does not produce definitive conclusions or statistical assessment of leaders' impacts. Moreover, there are many other countries that appear to experience no change in growth across leader deaths. Examples include Algeria, Israel, Taiwan, and Thailand (see Figure 2). In Taiwan, for example, the death of Chiang Kai-Shek in 1975, and the passage of power to his son, Chiang Ching-Kuo, appears to have been entirely seamless.[10] In the next section we pursue the question of whether leaders matter on average for economic growth using more rigorous econometric methods.

---

teachers and other governments workers who refused to accept pay cuts. Meanwhile, Cote d'Ivoire had to suspend and then restructure its debt payments in 1987.

[10] Note that a theory of strong leadership does not imply that a change in leadership *must always* coincide with a change in growth. In particular, one would expect no change in growth between the tenure of two leaders of similar quality and ideological bent.

3.3.2 Econometric Evidence

The key question in the following analysis is whether growth rates change in a statistically significant manner across randomly-timed leader deaths. Since we at the outset have no presumption that any particular leader is better or worse than any other, the test of whether leaders matter must consider whether growth episodes differ significantly before and after leader deaths, without presuming any particular direction for the change. Using our data, we will therefore estimate the following regression:

$$y_{ct} = \alpha_{cz} * PRE_z + \beta_{cz} * POST_z + v_c + v_t + \epsilon_{ct} \tag{3.1}$$

where y_{ct} is the annual growth rate of real purchasing-power-parity per capita GDP taken from the Penn World Tables, c indexes countries, t indexes time in years, and z indexes random leader deaths. Country and time fixed effects are included through v_c and v_t respectively. The PRE_z and $POST_z$ dummies capture deviations from country and time average growth rates before and after a specific leader's death. PRE_z is a dummy equal to 1 in the T years prior to leader z's death in that leader's country. $POST_z$ is a dummy equal to 1 in the T years after leader z's death in that leader's country. In the main analysis, we will choose the period of observation, T, to be five years, though in Section 6 we will show that the results are robust to choosing different periods. Note that PRE_z and $POST_z$ are defined so that the actual year of the death is not included in either dummy. This is probably the most conservative strategy when looking for longer-term leader effects, as it helps to exclude any immediate turbulence caused by the fact of leader transition itself.[11]

Figure 3 presents the results from estimating equation (1) with T equal to 5 years. The y-axis plots the estimated change in growth after the leader's death, $\beta_{cz} - \alpha_{cz}$, for each random leader death in the sample, and the x-axis plots the year of the random leader death. Figure 3

---

[11] The results in this paper are robust to a number of other methods of handling transition years. For example, assigning the transition year to either the PRE_z or $POST_z$ dummy, or assigning a fraction of the dummy to either the PRE_z or $POST_z$ dummy, produces similar or slightly stronger results than those presented here.

reveals that random leader deaths are associated with both increases and decreases in growth in approximately equal proportion—in 55% of cases, the estimated change in growth after the leader death, $\beta_{cz} - \alpha_{cz}$, is negative.

Under the null hypothesis that a particular leader z does not matter for growth, we expect that $\alpha_{cz} = \beta_{cz}$, i.e., that there is no systematic change in growth associated with the change in leader. That is, conditional on other regressors, we expect growth rates before and after the leader death to be similar. To answer the question of whether leaders matter for growth in general, we are interested in an F-test on the all leaders collectively. The null hypothesis is:

$$H_0 : \alpha_{cz} = \beta_{cz} \; for \; all \; c \; and \; z$$

If the error structure for ϵ_{ct} is *iid*, this F-test procedure will produce the correct inference.

However, we may be concerned both that the error ϵ_{ct} is neither identically distributed across countries nor independently distributed over time within the same country. In such cases, the F-test may not produce correct inference. To deal with these concerns, we will employ two strategies. First, we will attempt to determine the correct error structure and model the data generating process accordingly. Second, we will perform falsification exercises by running the F-tests on "control" regressions where the break points used to define PRE_z and $POST_z$ are moved several years backwards in time.

To investigate potential heteroskedasticity and serial correlation in growth, the first panel of Table 3 presents two tests for heteroskedasticity in growth across countries, a likelihood ratio test and a Breusch-Pagan Lagrange multiplier test. Both tests present strong evidence for heteroskedasticity in growth. This is not surprising: countries that concentrate their exports in a few industries, to take one example, are likely to have greater variance in their growth process. For the estimation strategy in this section, failing to correct for heteroskedasticity will have important implications for the F-test. To see this, consider that the estimators of interest, each α_{cz} and β_{cz}, are estimated primarily using within-country data.[12] If errors are heteroskedastic across countries, then assuming iid errors will overestimate the underlying variance in growth rates for some countries and underestimate it for others. In particular, if the correct variance for the data generating process in country c is σ_c^2 and the overall variance for all countries is σ^2, then the standard error on the coefficients α_{cz} and β_{cz} in that country will inflated by σ/σ_c. Since the individual restrictions in the F-test compare α_{cz} and β_{cz} within the same country, high (low) variance countries will show improperly high (low) significance for any difference of pre-post periods. This issue suggests that feasible generalized least squares with country-specific heteroskedasticity should be strongly favored over OLS estimation with a single σ^2.[13]

The second panel of Table 3 presents Breusch-Godfrey (Lagrange multiplier) tests for first-order autocorrelation on a country-by-country basis. Here we find evidence for autocorrelation in only 20% of the countries using a 95% confidence level. While this evidence for autocorrelation is not particularly strong, it may still be important to control for it: autocorrelation will tend to increase the probability that given time periods have higher or lower growth than average. This will overpower the F-test.[14] Again, we can employ feasible generalized least squares to estimate country-specific AR(1) processes. Note that the failure to find evidence of autocorrelation in most countries, together with the variation in the estimated autocorrelation parameter, suggest that country-specific autocorrelation is a better specification than assuming a single world-wide AR(1) process. As shown in Table 3, a likelihood ratio test strongly rejects a single world-wide AR(1) model in favor of country-specific autocorrelation.

---

[12] The only regressors in equation (1) common across country panels are time fixed effects.

[13] Another possibility would be to use White heteroskedasticity-consistent standard errors. However, as there are only 5 observations for each fixed effect, there are not enough observations for each variable to satisfy the consistency requirements of the White method. By estimating a single σ_c^2 across all observations for a single country, we have a much larger number of observations with which to estimate σ_c^2, and so the inference will be more accurate.

[14] Failing to account for serial correlation will overpower the F-tests, since the persistence of shocks implies that average growth during a given period is likely to be further from the mean than it would be if shocks were iid. Monte Carlo simulations conducted by the authors suggest that employing a GLS correction for AR(1) serial correlation corrects for this problem for the sample sizes used in this paper.

Table 4 presents the main results—i.e., the p-values from joint F-tests on the null hypothesis that $\alpha_{cz} = \beta_{cz}$ for all random leader deaths z. We present the results assuming four different specifications of the error structure. Column (1) presents OLS results. Column (2) presents FGLS results assuming country-specific heteroskedasticity. Column (3) repeats the specification in the second column but also allows for a worldwide AR(1) process, while column (4), the most general model, allows for country-specific AR(1) processes. Given the likely presence of country-specific heteroskedasticity and serial correlation, we focus on the results from the fourth specification.

For each specification of the error structure, we present three different timings of the PRE_z and $POST_z$ dummies as "treatments" and two different timings as "controls". The actual timing is represented by the row labeled t. On the theory that the effect of a leader's death may be felt with a lag, the timings $t+1$ and $t+2$ are included, indicating that the PRE_z and $POST_z$ dummies have been shifted 1 and 2 years later in time. The control experiments, using timings $t-5$ and $t-6$, repeat the hypothesis tests where the PRE_z and $POST_z$ dummies have been shifted 5 and 6 years backwards in time. Given that the timing of the leader's death is random, there should be no systematic events at $t-5$ or $t-6$ that produce shifts in growth regimes. These control experiments are included as evidence for whether the F-tests are overpowered.[15]

The results presented in Table 4 show that leaders have significant effects on growth. Using the contemporaneous leader timing (t), three of the four error specifications strongly reject the null hypothesis that leaders do not matter. Results are also strong for lagged leader effects, showing that we can again strongly reject the null when shifting the timing forward one or two years. Results are somewhat weaker when we assume a common AR(1) process across countries, but this specification received little support in our earlier tests (see Table 3). Importantly, the

---

[15] Note that we exclude Shastri of India, because his death came only 18 months after Nehru's death. Including both Shastri and Nehru will contaminate the other's PRE and $POST$ dummies in the regression. Given that Shastri's rule was far shorter than Nehru's, we have chosen to include Nehru in the main results, though all results are robust to including Shastri instead.

control experiments run at $t - 5$ and $t - 6$ do not show significant breaks in growth.

3.4 In What Context Do Leaders Matter?

The results in Section 3 show that, on average, leaders do matter for national growth in the developing world. However, the degree to which leaders matter will clearly be a function of their context. Different institutional systems and different social constraints can amplify or retard a leader's influence. If leaders do appear to matter on average, in what context do they matter the most?

A simple way to answer this question is to extend the above regression framework to consider hypothesis tests on subsets of the leader deaths. This approach preserves the non-directional nature of the exercise and allows us to consider the basic interaction of various national characteristics with the ability of leaders to influence national growth.

Where and when do leaders matter the most? Using the Polity IV data set (Marshall and Jaggers, 2000), we can compare the deaths of leaders by various institutional measures. In particular, the first panel of Table 5 compares those leaders whose nations receive Polity IV's lowest democracy score in the year prior to their death ("Autocrats") with those leaders whose nations receive better than the lowest democracy score ("Democrats"). The second panel in Table 5 employs an important sub-component of the democracy variable in Polity IV that defines the degree of institutional constraints placed on the national leader. The analysis divides leaders by whether the political system receives less than the median rating for institutional constraints in the year prior to death ("Unconstrained") with those polities ranked above the median ("Constrained"). The results indicate that autocratic and/or unconstrained leaders on average have a significant causative influence on national growth; meanwhile, those leaders in democratic or constrained regimes have little impact.

What is particularly interesting to know, given this result, is whether autocracy is a deterministic feature of nations, or whether all nations experience leaders who can change the national growth path. A simple tabulation of Polity IV's executive constraint measure shows that 65% of the 92 developing countries in this sample experience executive power both above and below the median value used to split the data in Table 5. Furthermore, 78% of the countries experience at least some period(s) of unconstrained executive power. These results suggest that powerful leaders can appear across a wide variety of settings.

Table 6 further explores whether influential leaders are produced (or not) according to deterministic national characteristics. One might expect that several explanatory variables used successfully in the recent growth literature might define those nations in which leaders do or do not have an impact. First, the well-known negative growth effect of being located in Sub-Saharan Africa suggests that we consider whether the leader results are a regional phenomenon. The first panel of Table 6 indicates that we do see strong leader effects in Sub-Saharan Africa; however, we also see strong effects in Latin America, which suggests that leaders' impact is not constrained to certain regions of the world.[16]

A similar set of results emerge when we consider ethnic fragmentation. Ethnic fragmentation is a strong negative predictor of growth (Easterly and Levine, 1997; Alesina et al, 2002) and helps predict institutional quality, including measures for the quality of government (La Porta et al, 1999) and corruption (Mauro, 1995). With regard to national leadership, one may imagine that ethnically fragmented nations provide particular opportunities for leaders to impact national outcomes. In this view, Tito and Milosevic could be seen as the difference between Balkan war and peace. Similar arguments could be made about the role of leadership in Nazi Germany, the recent case of Rwanda, and other cases of genocide.[17] The second panel in Table 6 indicates that leaders do matter in nations where ethnic fractionalization is high, but

---

[16] It is interesting to note that the OLS results suggest a much stronger impact in Sub-Saharan Africa than the FGLS results; conversely, Latin America shows no effect using OLS, but strong effects using FGLS. The reason for these differences is likely due to the underlying variance of the growth process in these regions. As discussed in Section 3.2, the homoskedasticity assumption in OLS will bias us towards finding stronger leader effects in highly volatile countries and weaker effects in less volatile countries. FGLS allows us to ask, more specifically, whether leaders matter in comparison to the growth experience of their nation alone. In an extremely strong leadership hypothesis, one could argue that leaders are all that matter to the variance in growth in a country, in which case OLS would be preferred. If we acknowledge that national volatility has other causes, then a heteroskedastic error structure appears preferable.

[17] Research has shown that ethnic fractionalization predicts internal warfare, although not when income is controlled for (Fearon and Laitin 2001).

they also matter where ethnic fractionalization is low.[18] Comparing the OLS and FGLS results, we see that there appears to be greater volatility in more fractionalized nations, resulting in extremely strong OLS results in those countries. Using FGLS to control for national volatility in growth, leaders are seen to have an impact irrespective of ethnic fractionalization.

British versus French colonization might be expected to predict where leaders matter, given the comparatively negative impact of French legal origin on property rights and democracy, among other institutional variables (La Porta et al, 1998; La Porta et al, 1999). One might expect, more specifically, that the British parliamentary system, which often results in coalition government, would lead to more constrained leaders than the French presidential system. Panel 3 of Table 6 investigates the differences between British and French colonial origin. The results are generally much weaker, suggesting that neither British nor French colonies have shown strong causative leader effects. Furthermore, while the comparison between these two cases is not definitive given the small sample sizes, the British versus French distinction does not appear to have any decisive influence here.

The final panel of Table 6 investigates the impact of settler mortality. Recent work has shown that the relative mortality of early colonial settlers is a strong predictor of current institutional quality (Acemoglu et al, 2001). Low settler mortality is argued in that research to have led colonial powers to settle extensively and import well-functioning European institutions with important limits on government power; high settler mortality led colonial powers to set up extractive regimes with strong central authority to transfer resources from the colony to the colonizer. In particular, high settler mortality is shown to predict weaker democracy and less constrained executives – the variables which succeed strongly in Table 5 in defining where leaders do and do not tend to matter. Not surprisingly, Table 6 shows that high settler mortality is a strong predictor of whether leaders matter, while low settler mortality shows no particular

---

[18] Leaders are grouped using a measure of ethnolinguistic fractionalization employed by Easterly & Levine (1997) that measures the probability two randomly selected individuals in a given country do not belong to the same ethnolinguistic group. In Table 6, "High Ethnic Fragmentation" indicates that a nation has greater than the median ethnic fractionalization measure; "Low Ethnic Fragmentation" indicates that a nation has less than the median fractionalization measure.

effects. These results suggest that regimes where leaders matter are predictable to some degree based on historical inheritances.

3.5 Are Autocrats Good or Bad for Growth?

The analysis in Section 4 showed that autocrats appear to matter for growth, whereas democrats do not. However, the analysis was purely non-directional—the F-tests used in Section 4 do not distinguish between whether the death of an autocrat led to an increase or a decrease in growth. This section examines the directional impact of leadership transitions.

To investigate this question, we employ a two-step procedure. In the first step, we estimate equation (1), from which we obtain an estimate of the change in growth after each leader transition. Using the notation of equation (1), the estimate of the change in growth after the death of leader z in country c is $\beta_{cz} - \alpha_{cz}$. In the second step, we estimate the following equation:

$$\beta_{cz} - \alpha_{cz} = \gamma_1 + X_{cz}\gamma_2 + Y_c\gamma_3 + \varepsilon_{cz} \tag{3.2}$$

where X_{cz} represent leader-specific characteristics and Y_c represent country-specific characteristics. We estimate equation (3.2) using weighted least squares, where the weights are equal to the inverse of the estimated variance of $\beta_{cz} - \alpha_{cz}$.

The results from estimating equation (2), where the dependent variable $\beta_{cz} - \alpha_{cz}$ is obtained by estimating equation (1) using 5-year pre and post periods, are presented in Table 7. The results show that the death of an autocrat is associated with an annual increase in growth rates of between 2.0 and 3.0 percentage points, relative to that which occurs following the death of a democrat. The results are even stronger when we use the more focused measure of executive constraints rather than the more general measure of autocracy. These results persist when including a host of other control variables, such as the leader's age and tenure in office in the year prior to his death, the country's degree of ethnic fragmentation, its colonial origin, and continent. Furthermore, none of these other explanatory variables appear to predict the direction of the change in growth following the death of a leader.

To further explore the impact of the death of autocrats, Tables 8a and 8b present a number of alternative specifications, in which we vary the timing and the length of the pre-and-post period used in estimating $\beta_{cz} - \alpha_{cz}$ and the controls used in estimating equation (2). Table 8a reports the results using the autocracy variable, and Table 8b reports the results using the unconstrained leader variable. Each cell reports the estimated coefficient on being an autocrat/unconstrained leader from a separate regression of the form reported in Table 7.

The results show strong evidence that autocracy matters, particularly when autocracy is measured using the lack of executive constraints variable. The results show that the differential impact between the death of an unconstrained and a constrained leader is still present even when we compare 7-year pre-and-post windows of observation. This suggests that eliminating a dictator produces prolonged changes in a country's growth pattern, rather than a temporary increase in growth due to the transition.

These results, which consider the leader as the causative force, also suggest a positive relationship between democratization and growth. The empirical literature on the effect of democracy on growth has produced ambiguous results (See Przeworski & Limongi (1993) for a survey), and exogenous sources of variation in political regimes have proved hard to find. More recent work has suggested that moves toward democracy are associated with higher growth rates (Minier, 1998) and that the presence of some democracy is better than complete autocracy (Barro, 1999). The results in this chapter are consistent with these recent findings.

3.6 Robustness of Results

The results presented above incorporate two kinds of robustness checks. First, they consider several different specifications for the error structure. Second, they consider control experiments as falsification exercises.[19] This section will consider two further types of robustness checks on the results: (i) the implications of different choices for the length of observation before and after leader deaths; (ii) the implications of conspiracy theories for the results.

In the main analysis presented above, average growth was compared for 5 year periods before and after each leader death. The choice of 5 years is essentially ad hoc and can only approximate the effect of leaders who may have been in power for substantially more or less than 5 years. As a general matter, we might think that choosing any fixed number of years for the comparison should bias the results against finding a growth effect, since we are capturing the actual tenure of the leaders poorly. In particular, a 5-year period may do too much justice to short term leaders and too little justice to long-term leaders, such as Mao, whose influence would have been felt over a much longer period.

One simple robustness check on the results is to consider observation periods of different lengths.[20] Table 9 reconsiders the growth regressions and hypothesis tests using a 3-year observation window and a 7-year observation window. The results appear essentially the same as those presented in Table 4. The results using a 7-year period are particularly strong, suggesting that the growth changes are quite persistent over time.

A separate issue is whether we should include short-term leaders at all. There are five leaders in the random leader sample who serve for less than one full year, and nine who serve for less than three years.[21] To take the most extreme example, Bouceif of Mauritania served for just over a month before he was killed in an air crash while trying to land during a sandstorm. It would seem unreasonable to imagine that this leader had an opportunity to influence Mauritania's growth given his short tenure. Introducing such short-term leaders may therefore bias the

---

[19] For clarity, the tables in Section 4 do not present results for control timings at $t-5$ or $t-6$. Those falsification exercises, which are available from the authors, consistently show appropriate degrees of noise.

[20] Using the actual tenure of leaders before and/or after the death is not an attractive option since it will introduce non-random events into the observation window (namely, the beginning of the initial leader's rule and the end of the latter leader's rule) which will prevent any causative interpretation of the results.

[21] There are 8 leaders all told in the random sample who served for less than 1 full year (see Table 2). However, three of these are already dropped in the regression; two are dropped because they die before the Penn World Tables begin to record data for their country; Ntaryamira of Burundi is dropped intentionally, as already discussed, due to the possibility his death was an assassination.

results against finding an effect. On the other hand, these leaders' short tenure implies that a non-random leader transition occurred relatively recently in the pre-period, which may in turn have been associated with some kind of change in the growth pattern. This could bias the results toward finding an effect. The second and third panels of Table 9 considers the sample of random leader deaths conditional on the leader having been in power for at least two, four, six, and twelve years respectively. We see that the results remain strong when short term leaders are dropped. Interestingly, the results seem to weaken when we look at the deaths of the longest tenure leaders – those in power for at least 12 years – though this may simply suggest that longer the leader was in power, the more inertia the regime has, resulting in more delayed growth effects. That interpretation is consistent with the shift we see in significance from t towards $t + 2$ as tenure increases in the table.

A more subjective issue is whether some random deaths might in fact be assassinations, and a definitive analysis of this issue is difficult, because conspiracy theories by their nature can be both hard to identify and difficult to refute. To address the possibility that some of the leader deaths were not random, the authors consulted biographical resources and examined large numbers of news articles in the year after each leader died, including obituaries and any follow-on articles about the death or ensuing effects in the leader's country. The results of this exercise show that no alternative theory for the death was presented in 45 of the 53 cases. At the same time, it is difficult to assess whether a "heart attack" was always as it seems, and a determined conspiracy theorist could assert than any number of heart attacks, air crashes, surgical deaths, or even cancers were part of a secret plot.

The authors have performed a series of robustness checks based on (i) confidence ratings for whether the death was truly random, and (ii) the type of death. In general, cutting the data to eliminate troublesome instances or types of death tends to make the F-test on the growth effects of all leaders less significant, and when all potentially troublesome instances or types of death are dropped, the F-test on growth effects often becomes insignificant. When the leaders are divided into Autocrats and Democrats, however, the Autocrats always show significance no matter how we attempt to cut the data. The fourth panel of Table 9 shows essentially the "worst case" results, where we eliminate any case where sources have mentioned any alternative story for the death, no matter how implausible.[22] Here we see that, while the results are weakened, the death of autocrats still leads to significant changes in growth.[23]

3.7 Conclusion

Recent work in the cross-country growth literature has suggested that growth in the typical developing country changes dramatically from one decade to the next. This observation suggests that growth is, to an important degree, a function of relatively short-run forces.

This chapter considers one possible force – the national leader – in explaining these growth experiences. Randomly-timed leader deaths are used as a natural experiment to identify the causative impact of leaders. We find that developing countries experience persistent changes in growth rates across these leadership transitions, suggesting that leaders have a large causative influence on the economic outcomes of their nations.

This chapter further shows that the effects of leaders are very strong in autocratic settings but weak or nonexistent in the presence of democratic institutions. Moreover, the deaths of autocrats lead on average to substantial, sustained improvements in growth rates. These results add texture to a growing literature on institutions in shaping economic outcomes. In particular, this research suggests that political institutions, separate from property rights or other institutional features, have large implications for growth. One interpretation of these results is that international intervention to remove autocrats may have a first-order economic

---

[22] This means, for example, excluding the plane crash that killed Samora Machel of Mozambique. His plane flew into a hillside at night while approaching the airport in Maputo. A conspiracy theory at the time suggested that the plane was lured off course using a false homing beacon across the border in South Africa. This hypothesis was later ruled out on several grounds, including the fact that other air traffic did not make this mistake. A more plausible conspiracy theory surrounds the death of Sani Abacha of Nigeria. As the *New York Times* reported, "Some United States intelligence analysts say there is evidence that Gen. Sani Abacha... was poisoned while in the company of three prostitutes... Nigeria's military rulers reported after General Abacha's death that he had died at his villa after a sudden heart attack. The contrary view reached by some United States Government analysts, while far from unanimous, is that he may have been killed by enemies in his notoriously corrupt and authoritarian military circle."

[23] Results using 7-year growth windows are considerably stronger than those presented.

basis. Of course, a leadership change caused by external forces may be very different from a natural leader death, and the policies used to effect such a change may have their own adverse consequences for growth.[24]

The authors' primary interest in this study is to improve our understanding of the forces behind economic outcomes. However, this research also informs a separate and very old literature in history and political science that considers the role of national leaders in shaping events. Deterministic views suggest that leaders have little or no influence, while the Great Man view of history, at the other extreme, sees history as the biographies of a small number of individuals. Tolstoy believed this debate methodologically impossible to settle, and prominent political scientists have more recently lamented a lack of scientific rigor in leadership studies (Paige, 1977; Blondel 1987). The leadership analysis in this chapter presents a methodology for analyzing the causative impact of leaders. We find that leaders do matter, and they matter to something as significant as national economic growth.

---

[24] Policy instruments that can promote leadership change include the leverage of international financial institutions, foreign aid, amnesty offers, economic sanctions, and military intervention. Such instruments are used often with leadership change in mind; recent examples include Robert Mugabe in Zimbabwe, Charles Taylor in Liberia, and Saddam Hussein in Iraq

Bibliography

[1] Acemoglu, Daron, Johnson, Simon, and Robinson, James A. "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, December 2001, 91, 1369-1401.

[2] Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. "Fractionalization," mimeo, Harvard University, 2002.

[3] Barro, Robert J. "Determinants of Democracy," *Journal of Political Economy*, 1999, 107, S158-183.

[4] Blondel, Jean. *Political Leadership*, London: Sage Publications, 1987.

[5] Easterly, William and Levine, Ross. "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, November 1997, 112, 1203-1250.

[6] Easterly, William, Kremer, Michael, Pritchett, Lant, and Summers, Lawrence H. "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks," *Journal of Monetary Economics*, 1993, 32, 459-483.

[7] Fair, Ray C., "The Effect of Economic Events on Votes for President," *The Review of Economics and Statistics*, May 1978, 159-173.

[8] Fearon, James and Laitin, David. "Ethnicity, Insurgency, and Civil War," mimeo, Stanford University, 2001.

[9] Gemmill, Gary, and Oakley, Judith. "Leadership: An Alienating Social Myth?" *Human Relations*, 1992, 45, 113-129.

[10] Marshall, Monty G., and Jaggers, Keith, Polity IV Project, Integrated Network for Societal Conflict Research Program and Center for International Development and Conflict Management, University of Maryland, 2000.

[11] Jerzmanowski, Michal. "Empirics of Hills, Plateaus, Mountains and Plains: A Markov-Switching Approach to Growth," mimeo, Brown University, 2002.

[12] Keegan, John, "Winston Churchill," *Time Magazine*, World Wide Web: http://www.time.com/time/time100/leaders/profile/churchill.html.

[13] La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, and Vishny, Robert W. "Law and Finance," *Journal of Political Economy*, 1998, 196, 1113-1155.

[14] La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, and Vishny, Robert W. "The Quality of Government," *Journal of Law, Economics, and Organizations*, 1999, 15, 222-279.

[15] Mauro, P. "Corruption and Growth," *Quarterly Journal of Economics*, 1995, 110, 681-712.

[16] Minier, Jenny A. "Democracy and Growth: Alternative Approaches," *Journal of Economic Growth*, 1998, 3, 241-266.

[17] Paige, Glenn D. *The Scientific Study of Political Leadership*, New York: The Free Press, 1977.

[18] Pritchett, Lant. "A Toy Collection, a Socialist Star and a Democratic Dud?" mimeo, Harvard University, 2001.

[19] Przeworski, Adam and Limongi, Fernando. "Political Regimes and Economic Growth," *Journal of Economic Perspectives*, 1993, 7(3), 51-69.

[20] Sachs, Jeffrey D. and Warner, Andrew M. "Fundamental Sources of Long-Run Growth," *American Economic Review*, May 1997, 87, 184-188.

[21] Tolstoy, Leo. *War and Peace*, 1869.

[22] Wolfers, Justin, "Are Voters Rational? Evidence from Gubernatorial Elections," Stanford GSB Working Paper #1730, 2002.

Major Data Sources:

[23] Bourgoin, Suzanne M. and Byers, Paula K. eds. *Encyclopedia of World Biography*, 2nd edition, 17 volumes, Detroit, MI: Gale Research Inc.,1998.

[24] Central Intelligence Agency. *The World Factbook 2002*, Internet Site: http://www.odci.gov/cia/publications/factbook/.

[25] Commire, A. and Klezmer, D. eds. *Historic World Leaders*, 5 volumes, Detroit, MI: Gale Research Inc., 1994.

[26] *Contemporary Black Biography*, periodical, Detroit, MI: Gale Research, Inc.1992-present.

[27] Frankel, Benjamin ed. *The Cold War: 1945-1991*, 3 volumes, Detroit, MI: Gale Research, 1992.

[28] Knight, Virginia C, ed. *African Biography*, 3 volumes, Detroit, MI: UXL, 1999.

[29] Lentz, Harris M. III. *Heads of States and Governments: A Worldwide Encyclopedia of Over 2,300 Leaders, 1945 through 1992*, Jefferson, NC: McFarland & Company, 1994.

[30] *The New York Times*, various issues.

[31] Rake, Alan. *100 Great Africans*, Metuchen, NJ: Scarecrow Press, 1994.

[32] Tenenbaum, Barbara A. ed. *Encyclopedia of Latin American History and Culture*, 5 volumes, New York, NY: Simon & Schuster, 1996.

[33] Zarate, Roberto Ortiz ed. *Zarate's Political Collections*, Internet Site: http://www.terra.es/personal2/monolith/home.htm, 2003.

Table 1: How Leaders Leave Power

92 Developing Countries
All Leaders from 1945 or National Independence Date through 2000
Number of Observations, by Type

Lost Election	Term Limits	Voluntary Retirement	Deposed	Death[a]	Other	Total
76	114	46	197	74	176	683[b]

Assassination	Natural	Accidental	
21	43	10	74

Heart disease	Cancer	Other Disease	Surgical complications	Other	Air crash	Other	
19	8	5	3	8	9	1	53

Notes

[a] There are 18 further cases (not included here) where leaders are killed during a coup.

[b] There are 771 distinct terms in which leaders are in power in the data set, but only 683 counted in this table, as we do not witness the exit of leaders who are still in power.

Table 2: Random Deaths of National Leaders

Country	Leader	Year of Death	Tenure	Nature of Death
Algeria	Houari Boumediene	1978	13.5	Waldenstrom's disease (blood disorder)
Angola	Agostinho Neto	1979	3.9	Cancer of the pancreas
Argentina	Juan Peron	1974	.7[a]	Heart and kidney failure
Bolivia	Rene Barrientos (Ortuna)	1969	2.7[a]	Helicopter crash
Botswana	Sir Seretse Khama	1980	13.8	Cancer of the stomach
Brazil	Arthur da Costa e Silva	1969	2.6	Paralytic stroke, then heart attack
Burundi	Cyprien Ntaryamira	1994	.2	Plane crash (assassination quite likely)
China	Mao Tse-tung	1976	26.9	Parkinson's disease
China	Deng Xiaoping	1997	19.2	Parkinson's disease
Comoros	Prince Sa'id Muhammad Jaffar	1975	.4	While on pilgrimage to Mecca
Comoros	Mohamad Taki	1998	2.7	Heart attack
Cote d'Ivoire	Felix Houphouet-Boigny	1993	33.3	Following surgery for prostate cancer
Ecuador	Jaime Roldos (Aguilera)	1981	1.8	Plane crash
Egypt	Gamal 'Abd an-Nasir (Nasser)	1970	15.9	Heart attack
Gabon	Leon Mba	1967	7.3	Cancer (in Paris)
Guinea	Sekou Toure	1984	25.5	Heart attack during surgery in Cleveland
Guyana	Linden Burnham	1985	19.2	During surgery
Guyana	Cheddi Jagan	1997	4.4	Heart attack a few weeks after heart surgery
India	Jawaharlal Nehru	1964	16.8	Stroke
India	Lal Bahadur Shastri	1966	1.6	Heart attack
Iran	Ayatollah Ruhollah Khomeini	1989	10.3	Following surgery to stem intestinal bleeding
Israel	Levi Eshkol	1969	5.7	Heart attack
Jordan	Hussein ibn Talal al-Hashimi	1999	46.5	Non-Hodgkin's lymphoma
Kenya	Jomo Kenyatta	1978	14.7	While sleeping
Liberia	William V.S. Tubman	1971	27.6	Complications surrounding surgery on prostate
Malaysia	Tun Abdul Razak	1976	5.3[a]	Leukemia (in London)
Mauritania	Ahmed Ould Bouceif	1979	.1	Plane crash in sandstorm over Atlantic
Morocco	Mohammed V	1961	5.3[a]	Following operation to remove growth in throat
Morocco	Hassan II	1999	38.4	Heart attack
Mozambique	Samora Machel	1986	11.3	Plane crash
Nepal	Tribhuvan	1955	4.1	Heart attack in Zurich
Nepal	Mahendra	1972	16.9	Heart attack
Nicaragua	Rene Schick Gutierrez	1966	3.3	Heart attack
Niger	Seyni Kountche	1987	13.6	Cancer (brain tumor)
Nigeria	Sani Abacha	1998	4.6	Heart attack (some say poisoned)
Pakistan	Mohammed Ali Jinnah	1948	1.1	Heart failure
Pakistan	Mohammed Zia Ul-Haq	1988	11.1	Plane crash
Panama	Domingo Diaz Arosemena	1949	.9	Heart attack
Panama	Omar Torrijos Herrera	1981	12.8	Plane crash
Philippines	Manuel Roxas y Acuna	1948	1.9	Heart attack
Philippines	Ramon Magsaysay	1957	3.2	Plane Crash
Rwanda	Juvenal Habyarimana	1994	20.8	Plane crash (assassination quite likely)
Sierra Leone	Sir Milton Margai	1964	3.0	After "brief illness"
South Africa	Johannes G. Strijdom	1958	3.7	Heart disease
Sri Lanka	Don Stephen Senanayake	1952	4.5	Thrown from horse
Swaziland	Sobhuza II	1982	60.7	Unknown
Syria	Abu Sulayman Hafiz al-Assad	2000	29.6	Heart attack
Taiwan	Chiang Kai-Shek	1975	25.3[a]	Heart attack
Taiwan	Chiang Ching-Kuo	1988	12.8	Heart attack
Thailand	Sarit Thanarat	1963	5.1	Heart and lung ailments
Uruguay	Tomas Berreta	1947	.4	During emergency surgery
Uruguay	Luis Ganattasio	1965	.9	Heart attack
Uruguay	Oscar Gestido	1967	.8	Heart attack

Notes: [a] Second time in power.

Table 3: Error Specification Tests

Panel 1: Cross-country specification tests

Null Hypothesis	Alternative Hypothesis	Test	Statistic	P-value
Homoskedastic	Heteroskedastic by country	LM (Breusch-Pagan)	Chi2(149)=3979.295	<0.0001
		LR	Chi2(149)=2981.43	<0.0001
Common AR(1) process	Country-specific AR(1) processes	LR	Chi2(149)=292.30	<0.0001

Panel 2: Breusch-Godfrey AR(1) tests by Country

			Percentage of countries that reject Null at given significance level		
Null Hypothesis	Alternative Hypothesis	Test	0.90	0.95	0.99
No serial correlation	AR(1) process	LM (Breusch-Godfrey)	26.7%	19.6%	7.1%

Table 4: Do Leaders Matter?

	P-values: Probability that average growth does not change systematically across randomly-timed leader deaths					
	(1) OLS	(2) FGLS	(3) FGLS	(4) FGLS	Number of Leader Deaths	Number of country-year observations
Treatment Timings						
t	.0328**	.0514*	.1462	.0582*	42	5544
t+1	.0345**	.0405**	.1239	.0676*	40	5544
t+2	.4073	.0371**	.0876*	.0444**	37	5544
Control Timings						
t-5	.8221	.5719	.7177	.6167	41	5544
t-6	.7238	.4915	.6864	.6796	39	5544
Country-specific heteroskedasticity	No	Yes	Yes	Yes		
Common AR(1) error	No	No	Yes	No		
Country-specific AR(1) error	No	No	No	Yes		

Notes:
(i) The table reports p-values, indicating the probability that the null hypothesis is false. Under the null hypothesis, growth is the same before and after randomly-timed leader transitions. Asterisks are used to indicate the significance with which the null is rejected:
* indicates 90% significance; ** indicates 95% significance; *** indicates 99% significance.
(ii) The regressions reported in this table compare 5-year growth averages before and after leader deaths.
(iii) All specifications exclude Rwanda and Burundi. Including these deaths will increase the strength of the results using treatment timings.

Table 5: Interactions with Type of Political Regime in Year Prior to Death

P-values: Probability that average growth does not change across randomly-timed leader deaths

	Autocrats			Democrats			
	(1) OLS	(2) FGLS	Number of Deaths	(3) OLS	(4) FGLS	Number of Deaths	Number of Obs
Treatment Timings							
t	.0005***	.0038***	23	.9860	.8466	19	5544
t+1	.0003***	.0015***	22	.9965	.9279	18	5544
t+2	.1636	.0043***	21	.7889	.6975	16	5544

	Unconstrained Leaders			Constrained Leaders			
	(5) OLS	(6) FGLS	Number of Deaths	(7) OLS	(8) FGLS	Number of Deaths	
Treatment Timings							
t	.0012***	.0062***	26	.9912	.7488	13	5544
t+1	.0008***	.0029***	24	.9886	.7951	13	5544
t+2	.1920	.0047***	23	.9992	.9571	11	5544

Notes:
(i) The table reports p-values, indicating the probability that the null hypothesis is false. Under the null hypothesis, growth is the same before and after randomly-timed leader transitions. Asterisks are used to indicate the significance with which the null is rejected: * indicates 90% significance; ** indicates 95% significance; *** indicates 99% significance.
(ii) FGLS regressions assume heteroskedasticity by country.
(iii) Distinctions across leader sets are defined using variables in the Polity IV data set in the year prior to the leader's death. Autocrats are defined by having the lowest score for the democracy variable (DEMOC). Democrats are those leaders with greater than the lowest score. Unconstrained leaders are defined by having the less than the median value of the executive constraint variable (XCONST). Constrained leaders are defined by having the median value or above.
(iv) To be conservative, all specifications exclude the 1994 plane crash that killed the leaders of Rwanda and Burundi.

Table 6: Interactions with Deterministic Variables

P-values: Probability that average growth does not change across randomly-timed leader deaths

	Sub-Saharan Africa			Latin America			
	(1) OLS	(2) FGLS	Number of Deaths	(3) OLS	(4) FGLS	Number of Deaths	Number of Obs
Treatment Timings							
t	.0001***	.0402**	14	.8163	.2001	10	5544
t+1	.0005***	.0884*	14	.3945	.0277**	10	5544
t+2	.2076	.4783	12	.2752	.0038***	9	5544

	British Colony			French Colony			
	(5) OLS	(6) FGLS	Number of Deaths	(7) OLS	(8) FGLS	Number of Deaths	
Treatment Timings							
t	.1377	.1708	14	.5887	.7789	10	5544
t+1	.1113	.2991	13	.8754	.9586	9	5544
t+2	.9968	.9779	11	.1391	.2234	8	5544

	High Settler Mortality			Low Settler Mortality			
	(9) OLS	(10) FGLS	Number of Deaths	(11) OLS	(12) FGLS	Number of Deaths	
Treatment Timings							
t	.0015***	.0758*	13	.9590	.1820	14	5544
t+1	.0025***	.0533*	12	.8921	.2345	14	5544
t+2	.2838	.0308**	11	.9867	.5911	13	5544

	High Ethnic Fragmentation			Low Ethnic Fragmentation			
	(13) OLS	(14) FGLS	Number of Deaths	(15) OLS	(16) FGLS	Number of Deaths	
Treatment Timings							
t	.0084***	.0866*	30	.7187	.1499	12	5544
t+1	.0093***	.0619*	29	.7185	.1604	11	5544
t+2	.4403	.1314	26	.3495	.0383**	11	5544

Notes:
(i) The table reports p-values, indicating the probability that the null hypothesis is false. Under the null hypothesis, growth is the same before and after randomly-timed leader transitions. Asterisks are used to indicate the significance with which the null is rejected: * indicates 90% significance; ** indicates 95% significance; *** indicates 99% significance.
(ii) FGLS regressions assume heteroskedasticity by country. Results are nearly identical when additionally allowing for country-specific AR(1) processes.
(iii) To be conservative, all specifications exclude the 1994 plane crash that killed the leaders of Rwanda and Burundi.

Table 7: How does growth change when different types of leaders die?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Autocrat	0.020**						0.030*	
	(0.010)						(0.015)	
Unconstrained Executive		0.029*** (0.010)						0.033** (0.016)
Age of Leader			0.00009				0.00082	0.00064
			(0.00037)				(0.00049)	(0.00046)
Tenure of Leader			0.00039				-0.00129	-0.00083
			(0.00062)				(0.00102)	(0.00089)
British Colony				-0.016			-0.012	-0.003
				(0.011)			(0.014)	(0.015)
French Colony				0.007			0.026	0.014
				(0.016)			(0.024)	(0.024)
Latin America					0.007		0.002	0.001
					(0.014)		(0.019)	(0.019)
Sub-Saharan Africa					-0.0002		-0.020	-0.017
					(0.0120)		(0.019)	(0.019)
High Ethnic Fragmentation						0.003 (0.011)	0.019 (0.015)	0.016 (0.014)
Constant	-0.01	-0.011*	-0.011	0.004	-0.001	-0.001	-0.057	-0.05
	(0.007)	(0.006)	(0.024)	(0.007)	(0.007)	(0.007)	(0.034)	(0.033)
Observations	39	39	39	42	42	38	36	36
R-squared	0.10	0.29	0.02	0.07	0.01	0.00	0.26	0.26

Notes:

This table presents the results from estimating equation (2) using weighted least-squares. The dependent variable is the average difference in annual growth rates between the five years after the leader's death and the five years before the leaders death, estimated by FGLS with country-specific heteroskedasticity using equation (1). Standard errors in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8a: Alternative specifications of the relative effect of eliminating autocrats

	Random Leaders Transitions		
	3 Year Windows	5 Year Windows	7 Year Windows
No Controls			
t	0.023*	0.020**	0.013
	(0.013)	(0.010)	(0.010)
t+1	0.025*	0.019*	0.017*
	(0.013)	(0.010)	(0.010)
t+2	0.014	0.014	0.020**
	(0.014)	(0.010)	(0.008)
Full Controls			
t	0.033	0.030*	0.025*
	(0.020)	(0.015)	(0.014)
t+1	0.036*	0.030*	0.029**
	(0.018)	(0.016)	(0.014)
t+2	0.016	0.024	0.027**
	(0.020)	(0.016)	(0.012)

Notes:
Each cell reports the coefficient on *autocrat* from a separate estimation of equation (2) using weighted least-squares. The regressions listed under "controls" also include the leader's age and tenure in the year before he died, colonial origin dummies, continent dummies, and a dummy for whether the country had a high degree of ethnic fragmentation as defined in Section 4. Number of observations varies from 32 to 39, depending on specification. Standard errors in parentheses.
* indicates 90% significance; ** indicates 95% significance; *** indicates 99% significance.

Table 8b: Alternative specifications of relative effect of eliminating unconstrained leaders

	3 Year Windows	5 Year Windows	7 Year Windows
No Controls			
t	0.032**	0.029***	0.025**
	(0.013)	(0.010)	(0.010)
t+1	0.035***	0.022**	0.021**
	(0.013)	(0.010)	(0.010)
t+2	0.022	0.012	0.018**
	(0.014)	(0.011)	(0.009)
Full Controls			
t	0.040*	0.033**	0.034**
	(0.020)	(0.016)	(0.015)
t+1	0.044**	0.029	0.029*
	(0.019)	(0.018)	(0.016)
t+2	0.03	0.023	0.023
	(0.021)	(0.018)	(0.014)

Notes:
See notes to Table 8a. Each cell reports the coefficient on *unconstrained leader* from a separate regression.

Table 9: Robustness Checks

P-values: Probability that average growth does not change across randomly-timed leader deaths

	3-Year Dummies			7-Year Dummies			
	(1) OLS	(2) FGLS	Number of Deaths	(3) OLS	(4) FGLS	Number of Deaths	Number of Obs
Treatment Timings							
t	.0440**	.0360**	42	.0081***	.0025***	42	5544
t+1	.4544	.0831*	40	.0078***	.0037***	40	5544
t+2	.7984	.0054***	39	.2480	.0519*	39	5544

	Tenure >=2			Tenure>=4			
	(5) OLS	(6) FGLS	Number of Deaths	(7) OLS	(8) FGLS	Number of Deaths	
Treatment Timings							
t	.0186**	.0220**	36	.0494**	.0672*	28	5544
t+1	.0164**	.0617*	34	.0241**	.0518*	26	5544
t+2	.5199	.0343**	31	.4524	.0536*	24	5544

	Tenure>=6			Tenure>=12			
	(9) OLS	(10) FGLS	Number of Deaths	(11) OLS	(12) FGLS	Number of Deaths	
Treatment Timings							
t	.2331	.1489	21	.9466	.7039	17	5544
t+1	.1034	.0509*	19	.7777	.3011	15	5544
t+2	.2622	.0356**	19	.6376	.0648*	15	5544

	Autocrats No Alternative Death Theory			Democrats No Alternative Death Theory			
	(13) OLS	(14) FGLS	Number of Deaths	(15) OLS	(16) FGLS	Number of Deaths	
Treatment Timings							
t	.0797*	.0580*	18	.9815	.8601	18	5544
t+1	.1160	.1171	17	.9943	.9144	17	5544
t+2	.5186	.2509	17	.7622	.7682	15	5544

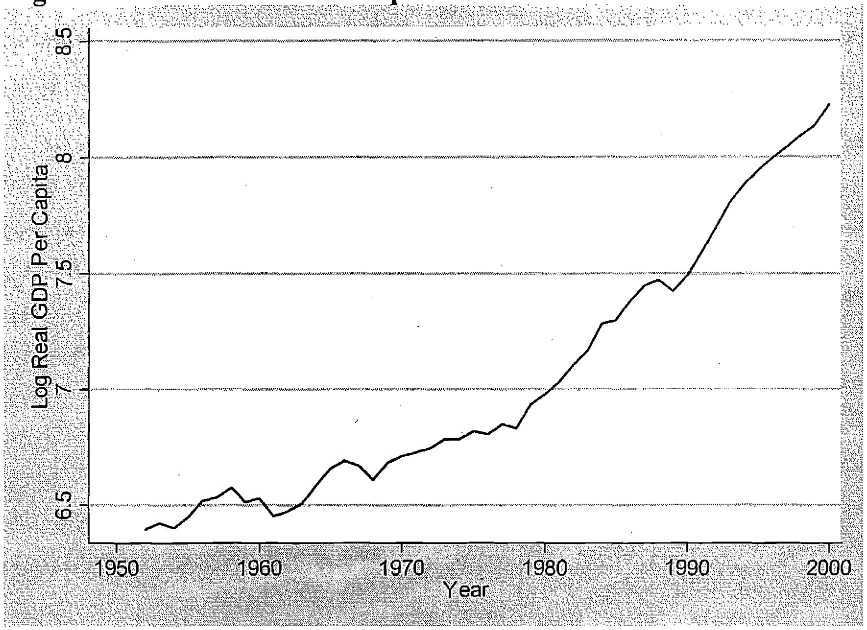
Notes:

(i) The table reports p-values, indicating the probability that the null hypothesis is false. Under the null hypothesis, growth is the same before and after randomly-timed leader transitions. Asterisks are used to indicate the significance with which the null is rejected: * indicates 90% significance; ** indicates 95% significance; *** indicates 99% significance.

(ii) FGLS regressions assume heteroskedasticity by country. Results are nearly identical when additionally allowing for country-specific AR(1) processes

(iii) To be conservative, all specifications exclude the 1994 plane crash that killed the leaders of Rwanda and Burundi.

Figure 1: China's Growth Experience



(source: Penn World Tables)

Figure 2: Growth and random leader deaths

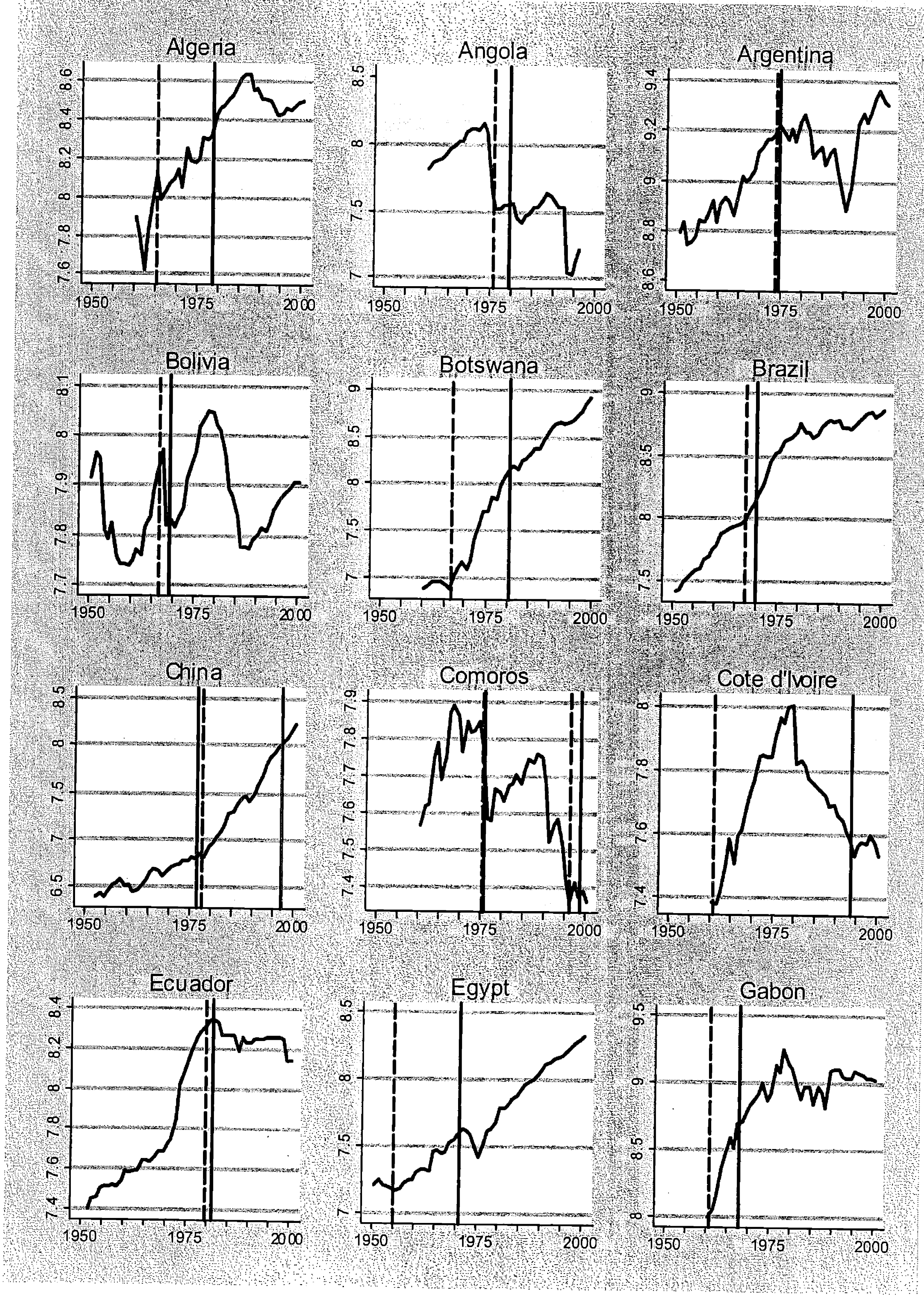


Figure 2 (continued)

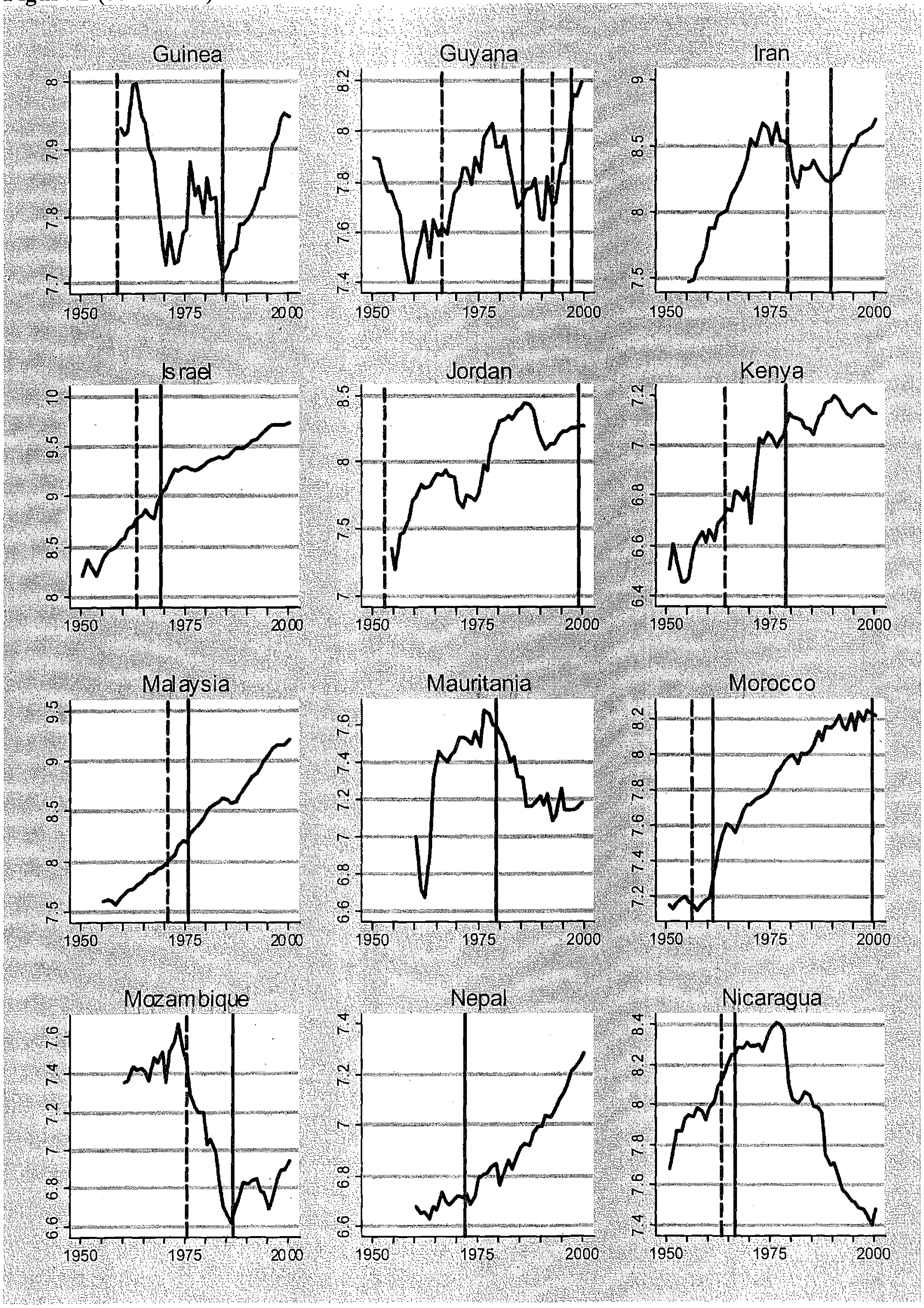


Figure 2 (continued)

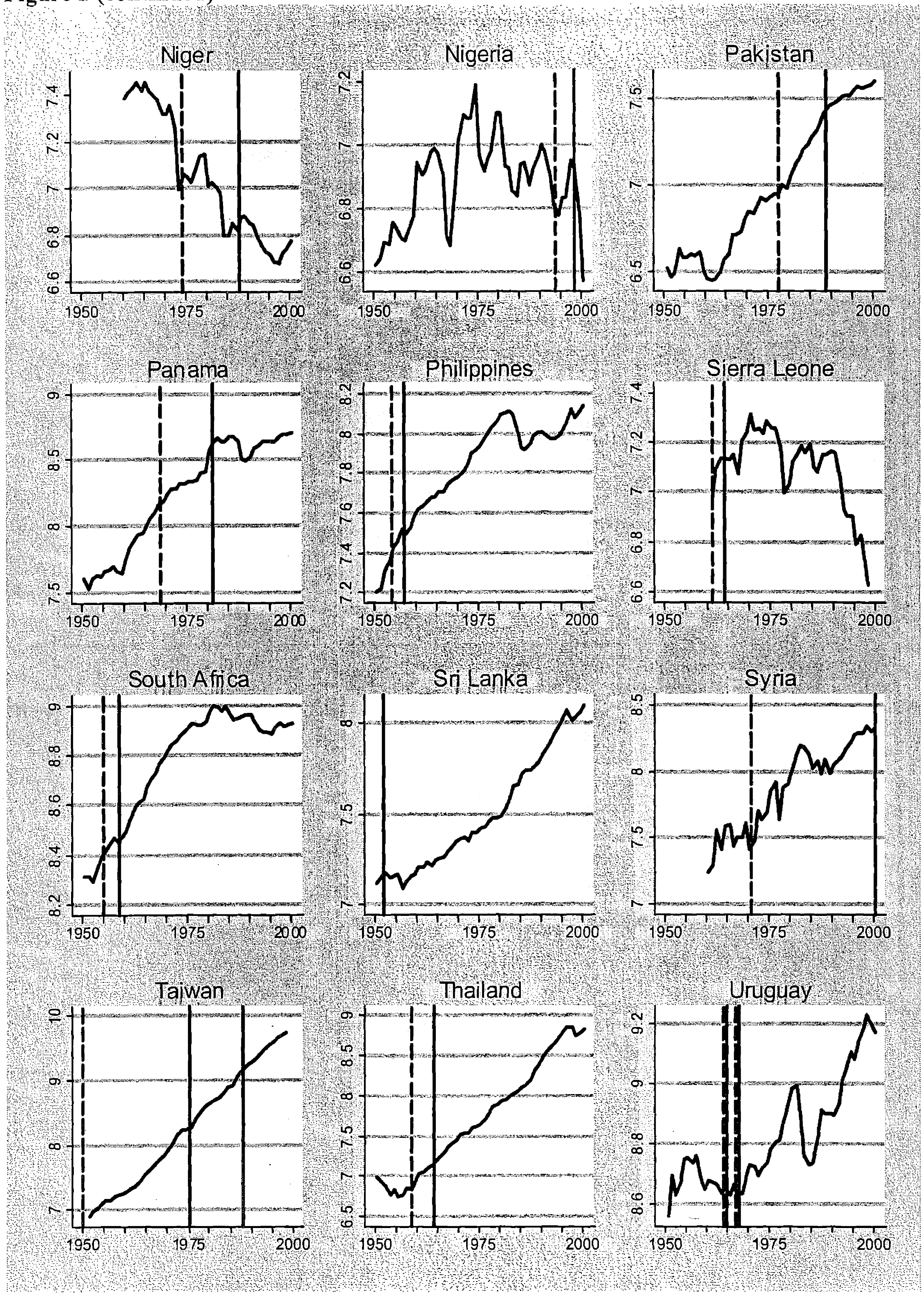


Figure 3: Changes in growth around random leader deaths