SCIENTISTS AND ENGINEERS IN ACADEMIC RESEARCH CENTERS—AN EXAMINATION OF CAREER PATTERNS AND PRODUCTIVITY

A Dissertation
Presented to
The Academic Faculty

By

James Scott Dietz

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Public Policy

Georgia Institute of Technology
February 2004

SCIENTISTS AND ENGINEERS IN ACADEMIC RESEARCH CENTERS—AN EXAMINATION OF CAREER PATTERNS AND PRODUCTIVITY

Approved by:

Dr. Barry Bozeman

Dr. Gordon Kingsley

Dr. Alan Porter

Dr. J. David Roessner

Dr. Sue V. Rosser

Date Approved: February 25, 2004

TABLE OF CONTENTS

DEDICATION V

ACKNOWLEDGEMENTS VI

LIST OF TABLES VII

LIST OF FIGURES VIII

LIST OF FIGURES VIII

GLOSSARY X

SUMMARY XI

Chapter 1: Introduction and Overview 1

Chapter 2: Basis in the Literature 7
2.1 Industrial Careers and Management of Innovation 7
2.2 Academic Careers and Productivity 9
2.3 Limitations of the Literature 21
2.4 Implications of the Literature for this Research 23

Chapter 3: Data Collection and Description 25
3.1 Data Collection Procedures 25
3.2 Description of Data 30

Chapter 4: Research Questions and Conjectures 38
4.1 Research Questions and Analyses 38
4.2 Constructs and Conjectures 39

Chapter 5: Data Analysis 42
5.1 Tobit Models of Publication and Patent Productivity 42
5.2 Poisson Model of Patent Count 52
5.3 Analysis of Publication and Patent Productivity Stars 56
5.4 Productivity Changes in Job Transformations to and from Industry 60
5.5 Neural Network Analysis 61

Chapter 6: Discussion of Findings 74
6.1 Publication Productivity 74
6.2 Patent Productivity 81
6.3 Similarities and Differences in Publication and Patent Productivity Analyses 85
6.4 Evidence in Support or Opposition to Hypotheses 87

Chapter 7: Limitations and Benefits of Using the CV for Career Research 92
7.1 General Issues in Using the CV as a Source of Career Data 92
7.2 Specific Problems with Certain Variables Included on CVs. 95
7.3 Benefits of the CV Approach 99

Chapter 8: Limitations of this Study and Future Research Directions 101
8.1 Sample Limitation 101
8.2 Limitations of Modeling and Model Assumptions 102
8.3 Validity of Measures 103

8.4 Limitation of Findings 105
8.5 Opportunities for Future Research 105

Chapter 9: Conclusions 109
9.1 Summary of Main Findings 110
9.2 Other Findings Relevant to Policy and Theory 116
9.3 Policy Implications of This Work 118

APPENDIX A: Tables 121

APPENDIX B: Tobit Model Diagnostics 158

REFERENCES 164

VITA 174

DEDICATION

For my mother Joan D. Dietz

and

For my father Louis P. Dietz

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation's Societal Dimensions of Engineering, Science, and Technology program and the U.S. Department of Energy's Office of Basic Energy Sciences. All findings and opinions do not necessarily reflect the opinions of these funding agencies.

I would like to acknowledge the support provided by the National Science Foundation. In particular, I would like to thank Jane Stutsman, Eric Hamilton, Barbara Olds, and all of my colleagues at the National Science Foundation who have lent their support and friendship, especially Giovanni Cruickshank and Barry Sloane.

I would especially like to thank my advisor, Barry Bozeman, for his wisdom, guidance, mentorship, and friendship, and for everything he has taught me and continues to teach me. The members of my committee—Gordon Kingsley, Alan Porter, Dave Roessner, and Sue Rosser—have provided invaluable support, mentorship, and advice. I would like to thank Phil Shapira for his ideas and interest in my work. I would like to thank both Alan Porter and Susan Cozzens for calming me down and firing me back up. In addition, there are so many others, too numerous to name, to whom I owe a debt of gratitude, including my family and friends.

LIST OF TABLES

Table 1. Center Titles and Response Rates 121

Table 2. Intercoder Reliability and Time of Coding for 10 CVs (Pilot Study) 125

Table 3. Intercoder Reliability for Coding 37 Items (Pilot Study) 126

Table 4. Descriptive Statistics 127

Table 5. Job Transformations Across and Within Sectors (Frequency and Relative Frequency) 129

Table 6. Tobit Model of Publication Rate (Number of Publications After the Year of the Doctorate as a Proportion of Career Length in Years) 130

Table 7. Mean Publication and Patent Rate, by Disciplinary Fields 133

Table 8. Tobit Model of Patent Rate (Number of Patents as a Proportion of Career Length in Years) 134

Table 9. Poisson Model of Patent Rate (Patent Count Divided by Career Length in Years) 137

Table 10. Differences in Means on Select Variables Between Publication "Stars" and Those Who Are Not Stars 138

Table 11. Differences in Means on Selected Variables Between Patent "Stars" and Those Who Are Not Stars 140

Table 12. Differences in Means on Selected Variables Between Publication and Patent "Stars" 142

Table 13. Differences in Means on Selected Variables Between Publication "Stars" and Those Who are Both Publication and Patent Stars 144

Table 14. Differences in Means on Selected Variables Between Patent "Stars" and Those Who are Both Publication and Patent Stars 146

Table 15. Differences in Means on Selected Variables Between Those Who are Both Publication and Patent "Stars" and those who are not stars. 148

Table 16. Mean Number of Publications For Five Years Before and After a Job Transformation from and to Industry (Standard Deviation, 95% Confidence Intervals) 150

Table 17. Neural Network Models, Estimated Output Values for Selected Input Values 151

Table 18. Summary of Tobit, Poisson, and Neural Network Models—Relationship Between Selected Independent Variables (Inputs) and Dependent Variables (Outputs) 156

Table 19. Summary of Differences in Means in Selected Variables Between Publication "Stars," Patent "Stars," Those Who are Both Publication and Patent Stars, and Those Who are Not Stars (Nonstars) 157

LIST OF FIGURES

Figure 1. Average Number of Publications Per Year by Career Age, All Respondents 33

Figure 2. Frequency of Total Number of Patents per Respondent 34

Figure 3. Frequency of Total Number of Jobs Held 35

Figure 4. Frequency of Total Number of Grants per Respondent (includes only those with one or more grants) 37

Figure 5. Classification of Stars 57

Figure 6. Neural Network Model: Relative Influence of Inputs in Modeling Publication Rate 65

Figure 7. Estimated Relationship Between Precocity and Publication Rate 66

Figure 8. Estimated Relationship Between the Number of Years in Government Jobs as a Proportion of Career Length 67

Figure 9. Estimated Relationship Between Industry Years as a Proportion of All Job Years and Publication Rate 68

Figure 10. Estimated Relationship Between Career Homogeny and Publication Rate 69

Figure 11. Neural Network Model: Relative Influence of Inputs in Modeling Patent Rate 70

Figure 12. Estimated Relationship Between Career Homogeny and Patent Rate 71

Figure 13. Estimated Relationship Between Years in Industry Jobs as a Proportion of Total Job Years 72

Figure 14. Estimated Relationship Between Precocity and Patent Rate 73

Figure 15. Estimated Relationship Between Industry Grants as a Proportion of Total Grants and Patent Rate 80

Figure 16. Summary of Findings on Postdoctoral Researchers Across All Models and Analyses 113

Figure 17. Summary of Career Homogeny Indexes Among Productivity Stars 115

Figure 18. Estimated Relationship Between Career Homogeny and Publication Rate 115

Figure 19. Estimated Relationship Between Career Homogeny and Publication Rate 116

Figure 20. Scatter Plot of Predicted Publication Rate by Error Squared (all cases) 159

Figure 21. Scatter Plot of Predicted Patent Rate by Error Squared (all cases) 159

Figure 22. Scatter Plot of Predicted Publication Rate by Error Squared (outliers above three standard deviations of the mean removed) 160

Figure 23. Scatter Plot of Predicted Publication Rate by Error Squared (outliers above three standard deviations of the mean removed) 160

Figure 24. Publication Rate Normality Diagnostic—Number of Observations 1, 2, 3, and More Than 3 Standard Deviations Above or Below the Mean Error 162

Figure 25. Patent Rate Normality Diagnostic—Number of Observations 1, 2, 3, and More Than 3 Standard Deviations Above or Below the Mean Error 162

GLOSSARY

CWRU	Case Western Reserve University
CV	Curriculum Vita
DOD	US Department of Defense
DOE	US Department of Energy
ERC	Engineering Research Center
GOALI	Grant Opportunities for Academic Liaison with Industry
HMM	Hidden Markov Model
IEEE	Institute of Electrical and Electronics Engineers
IFC	Interconnect Focus Center
IUCRC	Industry-University Cooperative Research Center
MD	Medical Doctor
MIRC	Microelectronics Research Center
MIT	Massachusetts Institute of Technology
NAS	National Academy of Science
NSB	National Science Board
NSF	National Science Foundation
USPTO	US Patent and Trademark Office
R&D	Research and Development
RVM	Research Value Mapping Program
S&T	Science and technology
SESTAT	Scientists and Engineers Statistical Data System
STC	Science and Technology Center
VPI&SU	Virginia Polytechnic Institute and State University
WebCASPAR	Computer-Aided Science Policy Analysis and Research system

SUMMARY

Science policymakers and research evaluators are increasingly focusing on alternative methods of assessing the public investment in science and engineering research. Over the course of the last 20 years, scientific and engineering research centers with ties to industry have become a permanent fixture of the academic research landscape. Yet, much of the research on the careers patterns and productivity of researchers has focused on scientists rather than engineers, specific job changes rather than the career as a whole, and publication productivity measures rather than patent outcomes. Moreover, much of the extant research on academic researchers has focused exclusively on the academic component of careers. As universities increasingly take on roles than were once considered the responsibility of the private sector—such as securing patents—and build greater ties with industry, it is timely to reexamine the nature of the contemporary "academic" career.

In this research, I draw on scientific and technical human capital theory to situate the central research question. Specifically, I examine the nature of the career pattern and publication and patent rates of scientists and engineers affiliated with federally-supported science and engineering research centers. The research makes use of curriculum vita (CV) data collected through the Research Value Mapping Program headquartered at the School of Public Policy. Tobit, Poisson, and Neural Network models are used in analyzing the data. In addition, I examine the career patterns of highly productive scholars and contrast those with less productive scholars.

The findings suggest that the ways in which academic productivity and career patterns have been conceived may be in need of revision, with a greater attention to diverse productivity outcomes and diverse career patterns. Some of the interpretations of empirical findings in the literature may be misconceived. Moreover, it may be the case that postdoctoral fellowship—a common component of government support for scientific and engineering research—may be associated with lower career productivity rates.

This research contributes to our understanding of research careers with implications for policies that may affect the outputs of governmentally supported research. Finally, the relatively

new method of collecting and analyzing CVs is discussed along with appropriate modeling techniques and the challenges posed by this method.

CHAPTER 1

INTRODUCTION AND OVERVIEW

Contributions to the scholarly knowledge base on the productivity of academic scientists and engineers have been considerable yet uneven in their conceptualization of the topic as a policy problem rather than a sociological issue, as a problem of career patterns rather than job promotions, and as problem situated in a scientific and technical social and human capital theory rather than one owed to a set of accumulating elite advantages. In this research, I examine the career patterns and productivity of a significant component of the contemporary academic research environment—scientists and engineers affiliated with university research centers that are expected to have significant ties with industry. This work is different than previous work in that careers are conceived of as a whole, both empirically and theoretically; both industrial and academic career experiences and outcomes are examined; and the methods used to collect and analyze the data are nearly untried and untested.

Historically, studies of the careers of scientists grew out of questions as to why there seemed to be such a lopsided distribution of research productivity across the population of academic scientists. As early as 1926, Alfred Lotka observed that the number of people producing "n" papers is k/n^2, where k is some constant.[1] In effect, a small minority of the population of scientists produces the vast share of published scientific work. But what explains the fact that many researchers have few publications over the lifetime, while others produce as many as 600 or more? For years, researchers in sociology, psychology, economics, and other disciplines have tried to explain why. Sociologists tended seek answers to this question in the sociological structures or "stratification" of science and the accumulation of advantages that accrue to elite researchers. Psychologists stressed innate factors, personality traits, and an inner sense of motivation. Economists have tended to take approaches rooted in human capital theory such as

[1] Price (1963) suggested that the square root of the population of scientists in any given field produce half of the scientific discoveries.

life cycle models. Most of these productivity and career studies have focused on academic to academic job changes as opposed to intersectoral careers, science as opposed to engineering, and publications rather than patents[2] as a productivity output indicator.

The largest body of work is sociological in nature and focuses on explaining publication productivity as a function of accumulative advantages that researchers receive, beginning with their affiliation with prestigious institutions of scholarly research which enable them to gain advantages not typically allotted to researchers in less prestigious institutions. Key variables have included the prestige of the doctoral department of researchers, their first job department, their current department, honorific awards, and peer judgments as to the quality of their work. Most of these studies focus on a single or several disciplinary areas and a single job change rather than the career as a whole. Researchers have also sought to explain productivity as a function of age and experience. Yet, these studies are less relevant to the formation of science policies and the evaluation of research.

The central question of this dissertation research is, "what effects do changes in jobs over the entire career have on productivity?" The main hypothesis to be tested, the "diversity" hypothesis, is derived from human and social capital theory and asserts that intersectoral changes in jobs throughout the career will result in higher productivity due to access to new social networks and human capital. Behind this assertion is the recognition that research is not an individually isolated phenomenon. It is socially and organizationally influenced—research ideas and questions and the methods used to pursue them depend upon a multitude of skills, perspectives, and collaborative thinking.

The main rival to the diversity hypothesis, the "homogeny" hypothesis, states that those who follow more "traditional" career paths will have higher productivity due mainly to differences in productivity incentives and disincentives across the three job sectors—academia, industry, and government. For example, academia prizes publication productivity, whereas industry may, at times, prohibit the open sharing of ideas and research findings.

[2] Incidentally, the largest body of work on patents tends to use the firm (or "assignee") as the unit of analysis.

Specifically, models of research productivity as measured in publication and patent rates of researchers are presented that include explanatory variables in several areas, including diversity of job experiences, grant awards, and early publication and mentoring opportunities. In contrast to the existing body of literature, many of these variables can be affected by the policies of funding agencies.

This approach differs from previous approaches in that it examines the pattern of researchers' careers over time and the effect of job changes and other critical events to the rate of productivity over time. It has its intellectual roots in a scientific and technical human capital (S&T human capital) theory (Bozeman, Dietz, and Gaughan, 2001) of knowledge generation that suggests that human and social capital building experiences over time affect the formation and pattern of scientific careers, and these opportunities intersect and act in synergistic ways to affect long-term productivity. The theory implies that a diversity of job experiences will affect collaborative patterns and the exchange of human capital through the building of a wider variety of network ties and social capital. This is similar to Granovetter's (1973) notion that greater benefits accrue to those who are able to tap "weak ties" (e.g., a friend of friend) due to exploitation of human and social capital that is non-redundant with one's one web of human and social capital endowments. The S&T human capital model itself will not be tested in this dissertation research because of the complexity of operationalizing the two critical components: human capital and social capital. However, the theory serves as a useful guide in exploring alternative and complementary approaches to the extant sociological, psychological, and economic literature.

This dissertation research is also methodologically different than previous research that takes purely correlational approaches (such as ordinary least squares regression) to explain variation that is explained in given a "dependent variable." While approaches such as these constitute a major portion of the analyses in this research, it is not the only approach. The logic of the S&T human capital model suggests that the career pattern itself (derived from curriculum vitae) is an equally appropriate unit of interest due to its broader landscape depiction of knowledge building over a lifetime. And, while difficult to operationalize empirically, the theory does serve to inform thinking about the opportunities that one receives in working in diverse research settings—

some of those opportunities are largely human capital in nature (i.e., cognitive knowledge, research craft knowledge, and other forms of tacit knowledge) or social capital in nature (i.e., the building of a more diverse collaborative network that permits human capital exchanges that are so vital to interdisciplinary research).

The proposed analyses include descriptive statistics of job changes, a comparison of the career patterns of publication and patent productivity "stars," Tobit models of publication and patent rates, and a Poisson model of patent counts. In addition, a preliminary set of Neural Network models will be used to analyze career patterns and their effects on productivity.

This research uses data from a US Department of Energy (DOE) and National Science Foundation (NSF) funded project called the *Research Value Mapping* (RVM) *Program*, which is headquartered at the Georgia Institute of Technology School of Public Policy. The program is studying new social and economic approaches to the valuing of publicly funded research—specifically as it is carried out in research centers. The curriculum vitae (CVs) of 1,200 research scientists and engineers supported by DOE, the Department of Defense, and NSF research centers have been collected and coded. In addition to the CV data, patent data were collected from the U.S. Patent and Trademark Office (USPTO) database.

It is expected that there will be several useful contributions from this dissertation research. First, this research will examine academic scientists and engineers many of whom have had prior experience working in government and industry. Traditionally, studies of scientific and technical careers have focused either on the industrial track or the academic track exclusively—one versus the other—as if careers were monotonic. Studies of industrial careers have their historical roots in the discipline of management and have focused on the management of innovation and the management of technical personnel, usually engineers. On the other hand, studies of academic researchers—having their roots in the sociology of science—have focused almost exclusively on the publication productivity of scientists (typically not engineers). The extant literature has not generally recognized the important knowledge synergies that may result from the human and social capital formation derived from job changes of multiple types across the sectors of academia, industry, and government.

The scientists and engineers in this study are affiliated with academic research centers that were designed to have industrial ties—a now seemingly permanent policy fixture of government support for science and engineering research. This provides an opportunity to test a model in an area where some form of commercial relevancy is expected and permits the examination of career factors that relate to academic, industrial, and governmental experiences. As universities take on more of the characteristics of the private sector, when it comes to the management and diffusion of intellectual property, and as industry continues to invest more money in academic research (from 3.8 percent of the total in 1980 to 6.9 percent in 2000 (NSB, 2002, Appendix 4-07)), this issue takes on greater policy relevance. In fact, for many years, the National Science Foundation has been encouraging greater interaction between the industrial and academic sector through programs like the Engineering Research Centers (ERCs), the Industry-University Cooperative Research Centers (IUCRCs), and the Science and Technology Centers (STCs) (these are incidentally the programs from which researcher career records were drawn for this study). And, more recently, the NSF has created a program called Grant Opportunities for Academic Liaison with Industry (GOALI) to encourage academic researchers to take sabbaticals in industrial research jobs and vice versa. It is timely then, from a policy point of view, to examine the knowledge impacts that these (nontraditional) careers may have.

Second, models of productivity and careers have not adequately addressed human and social capital factors, and there is growing recognition in science policy circles both in the U.S. (e.g., NAS, 1999) and Europe (e.g., Caracostas and Muldur, 1998) of the importance of these factors to international competitiveness and science policy in general. As of yet, however, few empirical studies have effectively examined science and technology policy as a form of human resources and human capital policy problem as the two research communities have remained largely separate. This research, although focusing on scientific careers, makes use of an interesting blend of variables related to research as well as to education and human resources.

Third, the data sources and analyses used in this dissertation are unique: the collection and coding of CVs and statistical modeling that is appropriate to the nature of these data. This dissertation research will serve as a methodological “proof of concept” for this approach and includes a discussion of its limitations and advantages and disadvantages.

Thus, I expect that this dissertation research will contribute to three broad areas of intellectual concern: (1) science policy (specifically research and development policy), (2) fundamental knowledge in the area of scientific careers and productivity, and (3) research evaluation techniques.

CHAPTER 2

BASIS IN THE LITERATURE

2.1 Industrial Careers and Management of Innovation

Traditionally, studies of scientific and technical careers have taken a narrow view, focusing either on the industrial track or the academic track exclusively. Studies of academic researchers (e.g., Keith and Babchuk, 1998)—having their roots in the sociology of science—have focused chiefly on the publication productivity of scientists and promotion in job rank. On the other hand, studies of industrial scientific and technical careers have their historic roots in the discipline of management and the management of innovation. They tend to focus on engineers (Goldberg and Shenhav, 1984; Allen and Katz, 1992), on the dual career ladder (Shepard, 1958; Allen and Katz, 1986; Gunz, 1980; 1989), on gatekeeping behavior (Turpin and Deville, 1995), innovation (Fusfeld, 1986; Burns, 1994; Rosenberg and Nelson, 1994; Mowery, 1998), technological obsolescence (Dalton and Thompson, 1971; Pazy, 1990; Bartel and Sicherman 1993; McCormick, 1995), and the management of technical personnel (e.g., Turpin and Deville, 1995; Debackere, Buyens, and Vandenbossche, 1997; Bowden, 1997).

In reality, however, many researchers change jobs between academia, industry, and government—sometimes changing sectors multiple times or working in multiple settings simultaneously. In the market context, economists often label this flow of knowledge from one organization to another “knowledge spillovers” (Jaffe, 1989; Griliches, 1992; Jaffe et al., 1993). In neoclassical economic thought, spillovers are considered to be inefficient to the operation of the market since creators of knowledge have difficulty capturing and containing its benefits. Yet, from a knowledge generation viewpoint, they can be viewed as often efficacious. The human and social capital that a researcher takes from one job to another (and perhaps from one job sector to another) may provide critical and ongoing knowledge inputs into new problems. The flow of people from one organization to another (such as from industry to academia and vice versa) is arguably key in the process of knowledge transfer, the diffusion of knowledge across organizations (Rogers, 1995), and the creation and maintenance of diverse knowledge networks over the career.

Yet there are few empirical studies that examine research ties between industrial and academic scientists and engineers. Zucker, Darby, and Armstrong (1998) examined geographically localized knowledge spillovers that occurred when superstar biotechnology researchers were affiliated both with universities and with firms. Through coauthorship patterns, they examined several types of human and social capital bonds between academic superstar researchers and scientists in local biotechnology firms, including collaborations (where there is no formal link between the two) and affiliations (where the superstars took on formal positions within the firm). What they found suggests this mixing of human and social capital may have very real economic returns. Five articles coauthored between the academic superstars in conjunction with firm scientists corresponded to five more products in development, 3.5 more products on the market, and 860 more employees for the biotech firms.

Landry, Traore, and Godin (1996)[3] provides one of the few studies seeking to relate research collaborations in the industrial setting with academic productivity[4]. They concluded that collaborative research in any form (between academic researchers and other academic researchers or between academic researchers and researchers in government and industry) “may indeed increase” the academic researcher’s productivity. They also found, however, that collaboration between academic and industry personnel had “significantly more impact” on productivity on the part of the academic researcher than did collaborations with other academics or government personnel.

---

[3] The findings of this study need to be approached with some caution since it is based on a survey questionnaire of academic researchers affiliated with universities only in Québec. The survey also suffered from a very low response rate (17 percent).

[4] Landry et al., formed a broad index of productivity as their dependent variable including the usual published outputs and patents but also lectures for nonscientific audiences, new courses, memoranda of expert opinion, supervision of graduate students, and other job-related duties of professors.

2.2 Academic Careers and Productivity

The lineage on academic research productivity begins principally in the 1950s with the notable work of Robert K. Merton (1957; 1961) to explain the social stratification of science as a sociological community, and with Anne Roe (1956; 1973) to explain the psychology of eminent scholars. But, it was not until the late 1960s and 1970s that academic productivity research blossomed with the works of Price (1963), the Coles (1967; 1970; 1973; 1979), Diana Crane (1965; 1969; 1970; 1972a), Paul Allison (Allison, and Stewart, 1974), and a host of others (Clemente, 1973; Faia, 1975; Reskin, 1977; Reskin, 1978; Friedkin, 1978). These scholars effectively set the agenda and the intellectual milieu for what was to follow. Some of these scholars were more concerned with the reward and recognition system of science, some with social structure and stratification, some with the distribution of resources and support, and others with more internally driven, innately defined factors.

Sociologists of science have long focused on scientific productivity in their studies of the sociological structures of science (Merton, 1961). These studies dealt with how science is organized as a sociological entity, not about how scientists and engineers produce and make use of knowledge. They had a profound and lasting effect on how careers in science are conceived of today. Their inadvertent effect, however, was to mark "science" as a province of the academic enterprise within its cultural norms (Merton) governing the creation, recognition, and sharing of knowledge. Industry, or industrial jobs within a largely academic career, was not considered in these studies.

Much of the scholarly work on academic productivity reflects these norms by placing social factors in the background, favoring instead measures of institutional and personal prestige, for example (Cole and Cole, 1967; Cole, 1970; Crane, 1970; Reskin, 1977; Long, 1978; Long, Allison, and McGinnis, 1979). Most of these models are essentially elite models that posit that scientific capital accrues disproportionately to those born of elite institutions with access to early advantages. A review of seven of the main approaches to explaining academic productivity of scientists follows.

2.2.1 Prestige Models

Prestige models are closely related to the accumulative advantage approach to explaining productivity (discussed in the next section). Essentially, using this approach, researchers examined the effects of institutional prestige on academic productivity. Among the first to do so was Crane (1965) who interviewed biologists, political scientists, and psychologists at three universities (a major private research university, a smaller university with some tradition of research, and a state university). Graduates of major universities were more likely to be highly productive than graduates of the minor university. She found that attendance at a major graduate school had more effect on later productivity than the current institutional location of the researcher, and that students of eminent sponsors were more likely to be highly productive later in their careers than other students.

Following Crane, Bayer and Folger (1966), using citation data as a productivity *qua* quality measure, studied 467 biochemists who earned doctorates in 1957 and 1958 and found a low but positive correlation between the prestige of the doctoral institutions and citations to their work published in 1964.

Later, Reskin (1977) examined the careers of a random sample of 238 doctoral chemists who obtained their Ph.D.s between 1955 and 1961. The caliber of the Ph.D. department exhibited a significant independent effect on the chemists' productivity at the end of their first professional decade. However, in contrast to Crane (1965), she found that beginning one's academic career in a university was a far more important determinant of decade productivity. Collaborating with one's sponsor during graduate school led to increased early postdoctoral productivity, but any direct effect was lost by the end of the decade.

Crane (1970) then examined the characteristics of faculty who joined the top 20 departments in six disciplines between 1963 and 1966. She found evidence that graduates of the highest ranking departments were much more likely to be hired by all the top 20 departments even after levels of productivity and citations were controlled thus giving credence to the prestige argument independent of productivity.

Allison and Long (1987), in investigating job changes of scientists, found weak but significant effects of productivity on the prestige of the destination job in 274 job changes made between 1961 and 1975 by academic physicists, chemists, mathematicians, and biologists. This contrasts with previous longitudinal studies that found no evidence of a relationship between research productivity and prestige of destination jobs. Major determinants of prestige of destination job included prestige of prior job, and the prestige of doctoral department. The number of articles published in the six years prior to the move had a lesser effect. Thus Allison and Long (1987) did identify a link from productivity to prestige of destination job although the effect size was essentially low.

Long, Allison, and McGinnis (1979) studied 239 male biochemists who earned their doctorates in 1957-58 and 1962-63. They found no statistically significant relationship between predoctoral publications and citations and prestige of first job, but they did find a relationship between prestige of first job and doctoral department prestige, number of citations to the mentor's work, and the selectivity of undergraduate institution. When researchers who had had a postdoctoral position were entered into the model, no relationship was identified between the prestige of first job and prestige of the postdoctoral institution. Although postdoctoral fellows had higher productivity, productivity still did not predict job prestige.

Allison and Long (1990) examined 179 academic job changes (between 1961 and 1975) of chemists, biologists, mathematicians, and physicists and found that publication and citation rates increased following a move to a more prestigious department and decreased following a move to a less prestigious department.

Long (1978) studied the effect of productivity on prestige of department for 134 male biochemists who had changed jobs and 47 who had not. He found that productivity increased for those who had moved to more prestigious departments, but only a weak effect of prior productivity on the move to the more prestigious departments.

Thus what can be gleaned from these early prestige studies is that institutional prestige—be it departmental or university—probably does have an effect on the prestige of destination jobs. More importantly however for this research, the link between prestige and productivity seems weak at best and perhaps spurious due to the caliber of doctoral advisers and mentors, increased opportunities to work with highly productive scholars, and the relative visibility of prestigious departments and those associated with them. This last study cited, Long (1978), provides a clue to the effects of job changing on productivity. Long found that productivity increased after a job change to a more prestigious institution, not before. This suggests that one aspect of this research, that job changing may increase productivity, is worthy of investigation.

2.2.2 Accumulative Advantage

The "accumulative advantage" hypothesis was first introduced by Merton (1968). With some convincing evidence, Merton and others (such as Cole and Cole, 1973; Allison & Stewart, 1974; and Allison, Long, and Krauze 1982) hypothesized that a succession of accumulating advantages bestowed on an elite group of budding scholars could explain why they tended to be more productive than researchers residing at second-tier institutions with fewer opportunities, less visibility, and fewer resources. Crane (1965), Bayer and Folger (1966), Reskin (1977), and Long, Allison, and McGinnis (1979) all found some positive relationship between the prestige of the doctoral department where the scientists had been educated and future productivity or job mobility. Other researchers (Cole and Cole, 1967; Cole, 1970; Long, 1978; Allison and Long, 1987; and Allison and Long, 1990) found the prestige of a previous academic job department to be positively related to the ability to move to more a prestigious job department and to subsequent productivity.

A further refinement of this prestige argument was what became to be known as the "accumulative advantage" hypothesis. Most of these studies dealt with the increased likelihood of citation to a researcher's work due to the accumulation of advantages. However, the hypothesis remains a plausible one in explaining publication productivity due to the increased resources and access to eminent scholars. Cole and Cole (1967) studied the scientific output of 120 eminent physicists and asked 1,300 additional physicists if they were familiar with the work of the eminent scholars and the honorific awards they had received. They found that the number of honorific awards, rank of the department, and percent of physicists familiar with the eminent scholars work correlated more highly with citation counts than with publication counts of the eminent scientists.

J. Cole (1970) studied 385 physicists who were highly cited in the Science Citation Index and found that scientists with high numbers of publications were more likely (51 percent) to cite authors in departments of distinguished rank compared with scientists with low numbers of publications (27 percent cited authors from distinguished departments). This is related to the so-called Matthew Effect, coined by Merton (1968) to refer to the status-enhancement effects of recognition. Derived from the Biblical phrase of the book of Matthew, "For unto every one that hath shall be given...but from him that hath not shall be taken away even that which he hath" (Matthew, 25: 29), Merton posited that the more one's work is cited, the more likely one's future work will be cited simply due to prestige factors.

Allison and Stewart (1974) used cross-sectional data to detect evidence for the accumulative advantage hypothesis. They examined citation patterns for 1,947 biologists, mathematicians, chemists, and physicists obtained through a probability sample and found a strong (positive) relationship between career age and the number of citations.

Allison, Long, and Krauze (1982) examined two sets of longitudinal data to test the accumulative advantages hypothesis: 239 chemists who earned their Ph.D.s between 1955 and 1961 and 557 male biochemists who received doctorates in 1957 or 1958 and 1962 or 1963. They found a strong tendency toward greater variation (and inequity) in publication counts among the scientists with increased professional age. The scientists' older publications were cited with less inequality (i.e., less variation in citation across scientists' works) than their more recent work.

In sum, there does appear to be something to the notion of the Matthew Effect in the citation to an eminent scholar's work. Like the prestige models, however, the accumulative advantages argument places emphasis on the prestige and structural advantages of being located at a highly visible department to explain the productivity increase. Equally plausible, however, is that the collaborative patterns and the exchange of social and human capital opportunities (even considering lower teaching course loads and more travel money) that allows these researchers to be productive.

2.2.3 Social Network Models

Some of the earliest work recognizing the importance of the social network to science and scientists was performed by Derek de Solla Price (1963) and Diana Crane (1968b; 1972b) on the concept of invisible colleges of scientists—roughly what Rogers and Bozeman (Rogers and

Bozeman, 2001; Bozeman and Rogers, 2002) call knowledge value collectives. Invisible colleges (Price, 1963; Crane, 1968b, 1972b; Carley, Hummon, and Harty, 1993; Hummon and Carley, 1993; Persson and Beckmann, 1995) are built on interpersonal relationships that facilitate various forms of collaboration (Katz and Martin, 1997) and communication among groups of scientists and permit them to exchange ideas and keep tabs on research within their own or adjacent fields. Invisible colleges depart conceptually from knowledge value collectives or other social network theories, in that they represent the "in-group" or prestige or power group within the field—the very core that those on the outside seek to emulate and who are enormously productive (Price and Beaver, 1966; Faust, 1997). The invisible colleges and the venues in which they operate—conferences, institutes, working groups, electronic communications—constitute both social inputs and outputs for individual scientists as well as science as a whole. This line of research recognizes that intellectual and scientific development occurs before, during, and after publication, and stresses that the all three are critical links in the knowledge chain (Merton, 1957b; Price, 1963).

Because of logistical complexities, empirical work on invisible colleges is rare, with few exceptions of note. Diana Crane (Crane, 1968; Crane, 1972) examined the communication patterns of rural sociologists in the diffusion of agricultural innovation. She traced direct and indirect ties in determining that a small group of productive scientists are directly interconnected with one another and attract an outer ring of less (or otherwise) productive scientists into indirect communication and influence. Mullins (1968) concludes from his comparison of communication ties among scientists that disciplinary orientation has obvious importance, but that scientists often communicate informally with scientists from other fields as well. Crane (1968b) suggests such influences are cultivated by scientists who wish to maintain exposure to new ideas in other areas of science. She found that scientists, in identifying others that influence their work, are just as likely to select scientists outside their disciplinary borders as within.

For Bozeman (Bozeman, Dietz, and Gaughan, 2001), scientific and technical human capital is embedded in social and professional networks or technological communities (Rappa and Debackere, 1992; Debackere and Rappa, 1994; Liyanage, 1995) of creators and users of knowledge called "knowledge value collectives." These networks integrate and shape scientific careers. They provide knowledge of scientists' and engineers' work activities, serve as resources for job opportunities and job mobility, and reveal possible applications for scientific and technical work products. Bozeman and Rogers define a knowledge value collective as a set of individuals connected by their uses of a body of scientific and technical knowledge[5]. It is a loosely coupled group of knowledge producers and users (e.g. scientists, manufacturers, lab technicians, students) pursuing a unifying knowledge goal (e.g. understanding the physical properties of superconducting materials) but to diverse ends (e.g. curiosity, application, product development, skills development) through the use and transformation of knowledge.

Despite the difficulties posed by social networks as a research method, what is most interesting about them for this research project, is that they place the productivity of individuals in relation to the productivity of the larger community of scientists working on similar problems. Due to the difficulty in operationalizing knowledge value collectives or social networks, this research focuses on federally funded centers as a proxy for at least one form (albeit an organizationally or interorganizationally delimited one) of social network.

What scholars like Price and Crane did not explore, however, is the nature and functioning of those links themselves. Best known for their consideration of this topic are Mark Granovetter (1973) and Ronald Burt (1992; 1997a; 1997b). Granovetter was interested in explaining how people get jobs through social networks and observed that they more often got them through distant social relations rather than proximate. He argued that "weak ties" (e.g., friend of a friend) represent social resources not available through stronger ties (e.g., family). People who have strong ties tend to share mutual friends and professional contacts; people with weak ties tend not to (Granovetter, 1973; Constant, Sproull, and Kiesler, 1996).

2.2.4 Innate Motivation

[5] For detailed treatment of the knowledge value collective and related concepts see Rogers and Bozeman (2001) and Bozeman and Rogers (2002).

From the realm of psychology, as early as the 1950s, Ann Roe (1953; 1956) was investigating the psychological characteristics of eminent scientists. She formulated a dizzying set of propositions concerning the personality traits of, for example, theoretical versus experimental physicists (all men). Roe (1953) consistently found that the eminent scientists she studied had an intense devotion to their work. Building on this work, Cole and Cole (1973) proposed the concept of the "sacred spark"—a largely intrinsically based argument of what motivates productive researchers. The Coles observed that many of the most productive scholars were driven by their own interest in the phenomena they studied and by the love of conducting research. However, and Bayer and Folger (1966) later noted, "…measures of intellectual ability or personality nearly always show very low correlations with productivity…It could be argued that adequate measures of scientific ability have not yet been developed, but this remains to be demonstrated" (p. 388). Through an early use of the Science Citation Index, they found that citation counts had a low but positive correlation with ratings of the quality of biochemists' graduate education, but no relation to measured I.Q. score.

2.2.5 Life Cycle and Life Course

Despite some good attempts, the extant literature has not managed to fully capture the dynamic nature of these career flows over time and across research contexts. Careers are inherently dynamic—evolving and intersecting in planned and unplanned ways, but traditional research evaluation models often view them as static or, at best, additive and cumulative over time.

Life cycle models attempt to address this problem by viewing the careers of scientists as a longitudinal function of the individual's skill levels and his or her incentives to act productively (Diamond, 1984; 1986). The concept originated in human capital theory from an economics tradition (Becker, 1964), which sought to relate investments in human beings (education, training, job and life experiences, and personal health) to an individual's earnings trajectory. As the argument goes, at earlier stages of career building, productivity incentives are strong while skills are growing. At the early to middle stages, both incentives and skills are strong as productivity peaks. And at middle to later stages, both wane, as does productivity (Levin and Stephan, 1991; Stephan and Levin, 1997; Simonton, 1997).

Clemente (1973) investigated the productivity of 2,205 Ph.D. sociologists who were members of the American Sociological Association in 1970 and found a negative correlation between age at first publication and later productivity, a weak negative relationship between age at Ph.D. and productivity, and a positive relationship between publication activity before receipt of the Ph.D. and subsequent productivity.

After Clemente's work, age[6] of the researcher came to dominate the work on productivity as exemplified by Diamond (1983; 1986), Simonton (1990; 1992), and Stephan and Levin (Levin and Stephan, 1989; Levin and Stephan, 1991; Stephan and Levin, 1997a). Cole (1979), through an examination of cross-sectional data, found a slight downward curvilinear relationship between age and publication output for 610 researchers in six fields of science. However, he also found, using data on 497 mathematicians who earned doctorates between 1947 and 1950, no drop off in productivity after 25 years. In fact, he found that the later work was more heavily cited than earlier work.

Bayer and Smart (1991) studied the career coauthorship patterns of 150 male chemists who received their Ph.D.s in 1960-62. They found that the proportion of single-authored papers regularly declines over the career, as does the proportion of dual-authored papers. Non-first-named authorship roles are frequent in the first two years of the post Ph.D. career, drop substantially after two years, and then increase substantially over the rest of the career. Multi-authored works dramatically increases over time, exceeding one-half of all papers for the most recent six-year career period.

Levin and Stephan (1996), who used longitudinal data from the Survey of Doctorate Recipients, found evidence in five of six scientific fields that scientists become less productive as they age. They conclude that research activity over the life cycle appears to be investment

[6] There is much debate about the relationship between age and research productivity. Some (Stephan and Levin, 1997) have observed a quadratic or logarithmic shape where productivity peaks in early to mid career. Others (Hunter and Kuh, 1987) refer to a "saddle shape" where productivity peaks twice. Still others suggest productivity can best be represented by a higher order polynomial (Simonton, 1990; 1992; 1997) with multiple productivity peaks.

motivated. Although there is plenty of empirical evidence to support this notion of diminishing marginal rates of productivity, there is some evidence to the contrary (Hunter and Kuh, 1987; Simonton, 1990; 1992; 1997) and such models have failed to explain much variation in productivity (Stephan, 1996). Moreover, as Stephan and Levin (1997) have pointed out, many of these life cycle models lack sufficient attention to the research process and the institutional setting of the process—something, incidentally, more akin to the concept of scientific and technical human capital.

Life course models can be thought of as an enhancement or conceptual expansion of life cycle models. The most important contribution of life course theories to the understanding of the scientific careers is the notion that human lives are linked or interdependent with other each other, and—not just statically—but dynamically over time. Merton ([1965] 1993) recognized this in titling his book, *On The Shoulders of Giants*, in which he illustrates how Newton made his intellectual advances using the contributions of his scientific peers and forefathers. Elaborated by Elder and Pavalko (1993) and Elder (1994), the life course paradigm views individual lives as affected by the historical period in which events occur, the developmental timing and sequence of events, and the involvement of the individual in relevant social relationships. Elder refers to the concept of human agency—which, as applied to science, can be thought of as the unique set of abilities that each scientist uses to translate his or her training and skills into scientific outputs. In a sense, human agency is a recognition that individuals vary in the predispositions (both strengths and weaknesses) they bring to the construction of a life course.

2.2.6 Education and Human Resources

Researchers have also called attention to the role of early career collaboration and mentoring as spurs to longer-term scientific productivity. Long and McGinnis (1985) found that predoctoral collaboration with mentors had significant and lasting effects on the careers of biochemists. The productivity of the mentor was positively and strongly related to the biochemists' own publication productivity six years later. For students who had not collaborated with their mentor, there was no relationship. Similarly, Reskin (1977), studying chemists who obtained their Ph.D. in the late 1950s, found graduates from higher “caliber” departments were more likely to have collaborated with their doctoral mentor and showed higher productivity after their first postdoctoral decade than graduates from lesser-prestige departments.

Zuckerman (1977) revealed that Nobel Prize winners viewed their doctoral apprenticeship as crucial to their later success. Specifically, the laureates pointed to its role in building broad skills such as knowledge of proper standards of achievement, tastes in choice of research problems, and confidence in their work and abilities.

Hunter and Kuh (1987), in summarizing the literature, suggest that socialization in a research environment is central to future productivity. The development of quality peer relations, the existence of a network of productive role models, the formation of a research orientation, experience as a graduate research assistant, and influence of team-based research learning activities all contribute positively to future career productivity.

2.2.7 Integrative Approaches

Stephan and Levin (1992) attempted to integrate the work of several research traditions by pointing to a utility-based view of scientific productivity where scientists pursue the extrinsic rewards of recognition and prestige among peers but also the intrinsic rewards of puzzle solving. They proposed three major motivators to productivity that interact with age over scientists’ careers: the puzzle (intrinsic pleasure), the ribbon (recognition), and the gold (rewards). They argued that three broad areas constitute something of a self-governing incentive system for scientists to behave in socially beneficial ways. Scientists continue investing in their own productivity until a point in the life course where further investments are unlikely to show “profitable” returns.

Most recently, there has been growing recognition in policy circles of a need to place more research emphasis on the socially-embedded nature of knowledge production (e.g., Callon, et al., 1991; Callon and Law, 1989; Callon, 1992; Courtial, Chlik, and Callon, 1994; Kostoff, 1994; Kostoff, Averch, and Chubin, 1994; Cozzens, et al., 1994; Sarewitz, 1996). In fact, as knowledge and information become more central to the functioning of the economy, they become more central to societal and social well-being and, thus, to policy. It could be argued that policymakers, in awarding grants for scientific research projects, should liken the process to putting bulbs in the ground—bulbs that grow scientific capacity over the long term even if they do not outwardly flourish in the immediate term. By favoring capacity—in the form principally of the generation of human and social capital—policymakers could be emphasizing policy-relevant variables that encompass not just knowledge outputs, economic outputs, or social outputs, but all three (Dietz, 2000). Arguably, then, public R&D evaluation should center not wholly on economic value or even improvements in state-of-the-art, but on the growth of capacity and S&T human capital (Bozeman and Rogers, 2002).

The S&T human capital approach to scientific productivity builds most prominently on the intellectual tradition of Crane's social network approach and others in the collaborative view of scientific productivity over the life course and Elder's work on life course theory. The S&T human capital perspective puts more weight on the human and social assets or endowments that scientists bring to their work than do previous models. Some are obvious like the formal assets, such as educational credentials and grant awards, that scientists and engineers accumulate. Other assets are more subtle, less formal, but perhaps even more important. Each scientist and engineer can be thought of as a unique embodiment of "S&T human capital"—a walking set of knowledge, skills, technical know-how and, just as important, a set of sustained network communications, often dense in pattern and international in scope (Dietz, et al., 2000). In previous work (Bozeman and Rogers, 2000; Dietz, 2000; Bozeman, Dietz, and Gaughan, 2001), we outlined an S&T human capital model as an alternative model for research evaluation, originating in response to the limitations of traditional economic, peer assessment, and case study approaches. S&T human capital includes not only the researcher's human capital but also the social capital he or she draws upon in creating knowledge and interacting in various social and professional contexts. It includes not just the educational credentials normally recognized in traditional human capital models (Mincer 1958; Becker, 1962; Schultz, 1963; Schultz, 1971; Coleman, 1993; Sweetland, 1996; Mincer 1997) but the researchers' tacit knowledge (Polanyi, 1967; Polanyi, 1969), craft knowledge, and know-how. And, essential to the effective exploitation of all of these human capital endowments is the social capital (Bourdieu, 1986; Bourdieu and Wacquant, 1992; Coleman, 1988; Coleman, 1990; Nordhaug, 1993; Putnam, 1993; Spagnolo, 1999) that scientists continually exercise in engaging their interests. These endowments not only make the study of scientists' and engineers' career trajectories more difficult than other professionals (e.g., less amenable to standard labor models) but more nuanced and more challenging. The movement of, and constant reshaping of S&T human capital is, arguably, a vital element of scientific discovery, technological innovation, and even economic development. As of yet, few studies have attempted to apply the concept of social capital to scientific research, some exceptions being Walker, Kogut, and Shan (1997); Gabbay and Zuckerman (1998); Nahapiet and Ghoshal (1998); Tsai and Ghoshal (1998); Bozeman, Dietz, and Gaughan (2001); and Dietz (2000).

2.3 Limitations of the Literature

Most of the literature on academic productivity stems from the 1960s and 1970s with less work in the 1980s and 1990s. Many of these studies tended to focus on a small number of, usually male, researchers and/or a small number of institutions. Moreover, studies with titles containing the word "career" are often about one job change—the first job after the doctorate, a change from a current position to the next, or a move to a higher job stature or status—and it effects on productivity. Few studies examine engineers, the effect of the proportion of one's career working in academia versus industry and government jobs, and patents as a productivity indicator.

Most often studies of patents and patenting involve cross-national comparisons (NSB, 2000; Radosevic and Auriol, 1999; Narin and Olivastro, 1998; Narin, Hamilton, and Olivastro, 1997; Graves and Langowitz, 1996), examinations of firm level behaviors (Klette and Griliches, 2000; Patel and Vega, 1999; Narin, 1995; Griliches, 1989; NSF, 1978), and are used in various studies of innovation and technological change (Kayl, 1999; Narin and Breitzman, 1995; Griliches, 1984; Pavitt, 1983, Schmookler, 1966).

In all, because of the limited choice of variables in the research (chiefly of interest to sociologists) and the inability to explain much variation in productivity, the utility of the body of literature remains of questionable value to policymakers. Thus, it may be that the way the problem has been framed has limited scholarly thinking about the links between productivity and human resources capacity building—something increasingly linked to science policy, economic development, and national innovation policies (Branscomb and Keller, 1998; Nelson, 1993; Nelson and Winter, 1982).

This, in turn, affects how problems and issues in science policy are conceived and the mental suppositions that inform research evaluation. The approach embodied in this study is different in this respect. To study the careers of scientists and engineers is to study how academia, industry, and government are often braided together through the knowledge embodied in one person over the life course.

The life course model, with all of its subtleties about the sequencing and timing of career events, seems to be more in line with the notion of scientific and technical human capital. And, thus, there are several important implications of scientific and technical human capital theory for the study of careers that span academia, industry and government. First, scientists, engineers, and technologists certainly do not exist in splendid isolation. They take part in various social institutions and organizations and much of their work is performed in that context. It does not make sense to reduce their capacity to create and use knowledge to one or two jobs or to the prestige of the organizations of which they are a part. Nor does it make sense to separate individual actors in the technical enterprise from their individual abilities and knowledge. Second, collaboration across organizational contexts (or the movement of personnel from one context to another) is a primary setting for the formation of fresh human capital through the interaction of its members and by the mutual tapping and flow of knowledge between them. Third, human lives are historical, ongoing, and linked. The paths and intersections of the lives of researchers affect their thinking, knowledge, and craft. The paths they choose have implications for the careers they build and for the scientific and technical human capital they create.

This is where S&T human capital theory is distinct from the accumulative advantages hypothesis, for example. S&T human capital puts more theoretical weight on the interpersonal social and human capital exchanges among researchers, as opposed to the accumulative advantages school of thought, which places more importance on a succession of job- and organization-related factors and prestige factors that promote or impede career attainment.

Finally, scientific and technical human capital facilitates the creation of greater capacity to generate knowledge and the diffusion of that knowledge among a wider community. It is in this way that study of careers can be viewed as one alternative to traditional models for the valuation of scientific investments.

2.4 Implications of the Literature for this Research

Taken as a whole, the extant literature suggests several factors that may be fertile ground for this study. Most of the studies on prestige and accumulative advantages suggest that productivity does not (or only weakly) predict the prestige of the next job department. Yet, when moving to a more prestigious (from a less prestigious) department, subsequent productivity does seem to increase. However, many of these studies use sampling frameworks that are tied to specific points in time, specific departments and institutions, or membership in specific disciplinary departments, making it difficult to duplicate on the scale and breadth of this study (which includes 5,490 specific job changes, starting in 1943). The RVM dataset is more diverse in terms of the scientists and engineering fields represented and the breadth of career length and institutional affiliation. Nevertheless, it does seem wise to make some attempt to control for institutional prestige. For example, research and development resource rich versus poor academic environments in which the scientist or engineer completed his or her doctoral education can be controlled as a way of examining the accumulative advantages hypothesis in the literature.

The social network approaches suggest that pattern and shape of social networks may be related to productivity and citation rates of scientists, but it is not clear if the network affects the productivity or the productivity affects the network membership. Unfortunately, social networks are known to be difficult to operationalize due to their exponentially increasing complexity as one maps out from ego and the difficulty in determining just where a network begins and ends. Nonetheless, the logic of the social network can at least be embodied through an examination of which of the centers included in this study are associated with relatively higher or lower levels of productivity.

Life cycle and life course approaches provide an important intellectual foundation at looking at careers as a whole over time. The life cycle and life course literature suggest that age affects productivity but this is in some dispute in terms of the nature of the relationship—most suggesting productivity is on the increase at early career ages and on the decrease in late career ages but it remains unclear what happens in the middle. Unfortunately, year of birth was recorded on only 34 percent of the CVs analyzed and it may be that this is correlated with those who got their doctorates in earlier decades when the practice of including personal information on the CV was more typical. Since productivity for the entire career will be examined for this dissertation, age becomes less important, especially if age-cohort effects are controlled for and productivity is examined as a rate (adjusted by career length).

The education and human resources approaches primarily suggest that working with strong and productive mentors, in a research-rich environment, and predoctoral publications will affect later productivity. This can be tested to some extent through examining the effects of predoctoral publications and postdoctoral research positions.

Finally, the integrative approaches give rise to much of the logic of this dissertation, including the central hypothesis that diverse job experiences allow for unique arrays of social capital ties and human capital endowments that yield higher productivity. This is not to say that prestige, accumulative advantages, social network, and life course approaches are incorrect. But it may mean that they account for only part of the picture. And, thus, it may be worthwhile to examine career sequences as complementary or additive to the effects found in the previous literature.

CHAPTER 3

DATA COLLECTION AND DESCRIPTION

3.1 Data Collection Procedures

3.1.1 Data Sources

This study makes use of a collection and coding of curriculum vitae (CVs) of scientists and engineers and an analysis of patent data from the US Patent and Trademark Office. Research associates who were part of the RVM program collected CVs from researchers affiliated with 90 NSF research centers, 2 DOE facilities, and two centers funded by the Department of Defense. We collected CVs from a total of 3,604 research scientists and engineers and graduate students working in these centers between January and October 2000. Each respondent received up to four email requests for his or her CV, which also contained a brief description of the study and assurances that the data would not be disclosed to third parties. 320 were dropped from the sample when correct email addresses could not be located. A total of 1,200 CVs were collected for an overall response rate of 36.5 percent[7] (see Table 1 for response rates by center). Because this study focuses on the careers of scientists and engineers only, CVs from undergraduate and graduate students and administrative support staff were deleted. The final number of CVs coded and analyzed for this study was 956.

Undergraduate, master's, and doctoral degree research assistants, who were supervised by two senior doctoral students, coded the CVs. Coders were introduced and trained one full day

[7] CVs were requested from scientists and engineers, students, and administrative support staff. The CVs from the latter two groups were not used in this study. Because it is difficult to determine what proportion of CVs requested were from students and administrative support staff, it is believed that the overall response rate for scientists and engineers exceeds 36.5 percent. Students and administrative support staff were less likely to believe that the email request was genuinely targeted at them. In addition, a number of responses from students were received stating that they did not have a CV.

in a web-based data entry form. When coders had technical questions they were directed to inquire to one of the doctoral student supervisors. The doctoral students kept a log of all coding problems and answers to frequently asked coding questions. The doctoral student supervisors extensively checked the data as they were being input into the system.

3.1.2 Pretest, Coding Procedures, and Intercoder Reliability Rates

Between October 1999 and February 2000, we conducted a pilot study in order to prepare for the NSF and DOE centers data collection. This pilot study included principal investigators from the NSF Biotechnology program, a collection of CVs from industry professionals working in biotechnology-related areas, a search of the internet for biotechnology-related CVs, and a collection of CVs from the Microelectronics Research Center (MIRC) at the Georgia Institute of Technology and its major inter-university collaborative research program, the Interconnect Focus Center (IFC) (these latter two centers were retained in the data analysis for this study).

The pilot study was completed in March 2000 and, even though not fully used in this analysis, it is perhaps useful to review what was learned from this study. For the NSF, industry, and research center collections, an email message was sent directly to potential respondents who were asked to submit a full CV via email. In contrast, for the Internet search, various search engines and search phrases were tested to identify a subgroup of web-posted CVs. Of the sample group, 50 CVs were solicited from industry scientists and engineers, 200 from NSF-funded academic researchers, 100 from the web, and all faculty and graduate students affiliated with the multi-institutional research center and its primary research program (n=210).

To develop a preliminary coding protocol, a subset of the CVs was reviewed from each of the four respondent groups to identify problems and potential solutions. We identified over 30 potentially useful variable “sets.” However, many of these variable sets included multiple (i.e., up to 10) degrees received, multiple (i.e., up to 600) publications, and so forth. The number of variables for each respondent depends, unlike a questionnaire, on the length of the CV. Junior researchers could have as few as 25 variables per CV; seasoned veterans could have as many as 3,000. Several practice coding exercises were conducted to obtain information on intercoder reliability, to improve the coding protocol and process, and to minimize coding time. After these preliminary steps, we revised the coding protocol, retrained the coders, and proceeded to code the 281 CVs collected from the four respondent groups. The goal was to design the coding process so that an undergraduate student could be trained to code the typical CV with minimal reliability problems in 30 minutes or less.

To test intercoder reliability, the work of five coders on a subset of 37 variables from two sets of 10 CVs was examined. We used Crittenden and Hill's (1971) measure of intercoder reliability (Rs) to test for intercoder reliability. Tables 2 and 3 summarize the results from the first preliminary coding tests. Overall, the average reliability coefficient value of .766 on the first round of coding suggested that further refinement of the protocol and coding scheme was needed. While there is no widely accepted "threshold level" of intercoder reliability, a coefficient below .850 should be considered problematic; a coefficient below .600 is regarded as unacceptable. Only 16 out of 37 items satisfied the .850 requirement. Moreover, 7 out of 37 items fell below .600. The principal coding problems stemmed from the limited standardization in CV formats, missing information, and coder error or the misinterpretation of data.

A closer inspection of the errors, however, demonstrated that many of them were due to coders coding information out of order (e.g., coding the second publication in the third publication variable spot), which caused a succession of errors compounding the problem and depressing intercoder reliability (even if the coding was technically accurate after the initial error occurred). This problem in effect served to artificially lower the intercoder reliability rates because coding problems such as these are not errors per se and do not affect data quality if the data are reordered after the fact. In addition, coders had significant problems with the original codebook, which was revised and resubmitted to a second test (which is labeled coding trial 2 in Tables 2 and 3) using the same coders but new CVs. We expected there would be improvements due to the enhanced codebook as well as learning effects, and we were correct. Tables 2 and 3 show that the mean intercoder reliability rate increased to .805. There were 15 items that scored above the .850 level and three that scored below .600. While, in general, these intercoder reliability rates may be viewed as on the low side, we expected rates lower than those achieved in more typical questionnaire coding due to the complex nature of the coding task. In addition, the actual intercoder reliability is quite a bit higher because the rates reported include errors of chronological ordering of data elements. But to address the problem directly, a more elaborate coder training program was put in place and coders were more closely supervised than originally expected. Finally, to ensure high quality data, one of the senior doctoral students was put in charge of periodic data checking and ongoing retraining of the coders when errors were detected.

3.1.3 Data Cleaning

Because of the complexity of the coding task, I undertook an extensive project of cleaning the data. For any case where there was extensive missing information or anomalous data (data out of range), the CV was retrieved and reexamined. The dataset was corrected when problems were discovered. I also analyzed records for each coder who was discovered to have made frequent errors and made appropriate corrections to the dataset. In many cases where there were extensive missing data or where the CV appeared to be abbreviated, I searched the relevant website for a more complete and fully annotated CV. The records were updated appropriately. Variables where high coding errors (due to task complexity) were detected (e.g., job type variables) were rechecked for accuracy. Finally, I resorted all records to correct for chronological coding errors detected in the pilot study. The final dataset contains approximately 3400 variables.

3.1.4 Patent Data Collection

I collected data from the US Patent and Trademark Office (USPTO) online database containing information about patents from 1976 to the present. Patents granted prior to 1976 were retrieved from the CV (although it is acknowledged that this is a less reliable source). These data contain information about other patents that are cited in the applications as well as scientific and technical publications that are cited. The data provide a description of the patent and its claims, the inventor(s) and assignee(s). For each of the 956 scientists and engineers, the database was searched for patent information. I tried several variations on each researcher's name (e.g., John

Doe, John S. Doe, John Scott Doe, J. Doe, J.S. Doe, and Doe) until the correct records were located. The year of the patent application and the institutional assignee were matched to the respondent's CV job record to ensure the correct person was identified.

The resulting dataset, which includes the CV data as well as patent data, is referred to as the RVM dataset in this dissertation study.

3.1.5 Project Support and Background

This research is supported through National Science Foundation and the US Department of Energy grants awarded to Barry Bozeman and Juan Rogers under the *Research Value Mapping* (RVM) Program within the School of Public Policy at the Georgia Institute of Technology. The RVM Program began in 1996 (Kingsley, Bozeman, and Coker, 1996) using 30 intensive case studies of research projects as sources of both qualitative and quantitative information about the nature and intensity of the projects' scientific and socioeconomic impacts. The Phase I[8] work, sponsored by the US Department of Energy's (DOE) Office of Science, focused entirely on DOE-sponsored projects in government and university labs. Phase II is based on continued funding from DOE with new funding from the National Science Foundation (NSF) and focused on knowledge impacts using the curriculum vitae of scientists and engineers primarily affiliated with NSF-funded centers. The Phase II population is made up of researchers affiliated with NSF science and technology centers, engineering research centers, and industry-university cooperative research centers and several Department of Energy and Department of Defense centers. This dissertation research is part of Phase II.

[8] For more information on RVM Phase I, see Bozeman, et al. (1997; 1998; 1999).

3.2 Description of Data

3.2.1 Sample

The RVM Program data collection was designed to represent projects funded by NSF and DOE and primarily reflects their need to understand the projects they fund. The population is made up of scientists and engineers working at federally funded research centers when the data were collected in 2000.

However, because many of the research centers included in this study are purposively interdisciplinary and problem driven in mission, it is expected that the data more heavily represent researchers working in interdisciplinary fields of science and engineering. It is also plausible to suspect that the researchers are also more likely to have had industrial jobs than other academic researchers (due to the focus of the NSF and DOD centers in particular on building bridges to industry) and with more overall external grant support (again because they are part of government funded centers) than the academic population in general. Finally, many of the centers (although by no means all) are located at elite research universities. So, although the data cannot be generalized to all scientists and engineers in all fields, they may be more relevant to policymakers at these important funding agencies.

3.2.2 Demographic Information

Approximately 12 percent of the respondents in the RVM dataset are women and 87 percent are men (see Table 4 for descriptive statistics on all major variables included in this study). This indicates that men are somewhat overrepresented in the RVM dataset. According to NSF, approximately 20 percent of all doctoral natural[9] scientists and engineers in 2001 were women (NSF 2001, Appendix Table 5).

[9] This figure excludes social and behavioral scientists since there are few in the RVM dataset and were not the focus of the RVM study.

Year of birth was indicated on 34 percent of the CVs. For the CVs that did contain this information, the range was 1923 to 1974 and the mean was 1950. Race and ethnicity information is typically not recorded on CVs and was not collected for this study. In terms of international scientists and engineers, approximately, 13 percent of recipients received their doctorates from foreign institutions and 27 percent received their bachelor's degree from a foreign institution. To put this in perspective, NSF reports that in 1999 approximately 20 percent of academic tenure or non-tenure track faculty scientists and engineers (in all fields) were non-native born (SESTAT, 1999). In engineering, the proportion was 34 percent (NSF, 1999).

3.2.3 Educational Background

In terms of disciplinary field differences, approximately 45 percent of the respondents in the RVM dataset earned their doctorates in engineering, 28 percent in the physical and mathematical sciences, 11 percent in the biological and agricultural sciences, 5 percent in the social and behavioral sciences, 5 percent in computer and information science, and 2 percent in the medical or health sciences (including medical doctor degrees (M.D.s)). For approximately 5 percent, the field of doctorate was missing.

In contrast, according to the National Science Foundation *Survey of Earned Doctorates 1999*, of all doctorate scientists and engineers working in academic institutions in the U.S., approximately 11 percent in earned their doctorates in engineering, 22 percent in the physical and mathematical sciences, 30 percent in the biological and agricultural sciences, 32 percent earned their doctorates in the social and behavioral sciences, 2 percent in computer and information science, and 2 percent in the medical or health sciences (NSF 2002, Appendix Table 19). Thus, the RVM dataset has a substantially higher proportion of engineers and physical and mathematical scientists and a lower proportion of biological and social and behavioral scientists than the overall US doctoral scientist and engineering academic workforce.

3.2.4 Career Age and Cohorts

The length of the respondents' careers (measured in years since doctorate or equivalent (e.g. M.D.)) ranged from 1 to 58 with a mean of 18 years (standard deviation 11.5). To test for effects of institutional or other educational changes over time, the respondents were divided into quintiles. Roughly one-fifth received their doctorate before 1972, one-fifth between 1972 and 1980, one-fifth between 1981-1987, one-fifth between 1988 and 1993, and one-fifth between 1994 and 2000.

3.2.5 Publications and Patents

The total number of publications per respondent ranged from 0 to 628 with a mean of 75 (standard deviation 92) and a median of 44. The average number of publications per year ranged from 0 to 34. Figure 1 shows that by seven years after the doctorate, the average publication rate was 2.8 per year. The distribution of publications across scientist and engineers is skewed (skewness[10] value of 2.59 for the total number of publications; 2.49 for the number of publications as adjusted by career length in years), suggesting that there are important extreme cases that make the right tail of the distribution of publications longer than is the case with the normal distribution.

[10] Skewness is a measure of the symmetry or asymmetry of a distribution. It takes on a value of zero when the distribution is normal. A high positive value indicates a long right tail, which is the case with the publications variable.

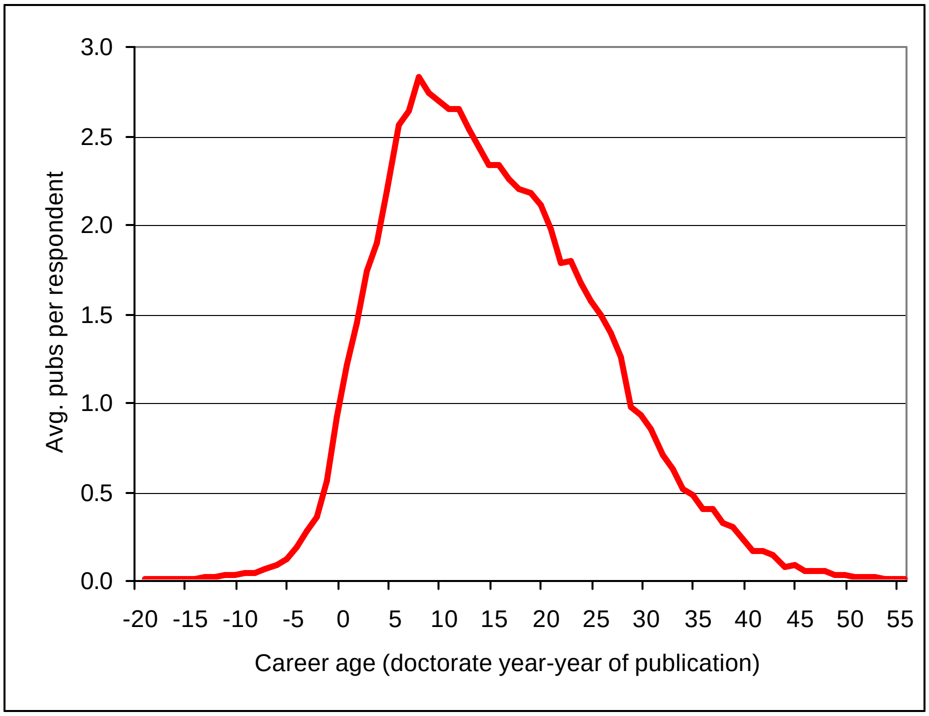

Figure 1. Average Number of Publications Per Year by Career Age, All Respondents

The total number of patents per respondent ranged from 0 to 141 with a mean of 2.7 patents per respondent (standard deviation 8.2) and a median of zero. Approximately 41 percent of respondents had been granted one or more patent. As Figure 2 shows, the distribution of patents among scientists and engineers is highly skewed.[11] Again this suggests non-normality and a long right tail in the distribution of patents among scientists and engineers.

---

[11] Skewness value of 9.02 for total number of patents; and a value of 6.68 for number of patents granted as a function of career length (i.e., patent rate).

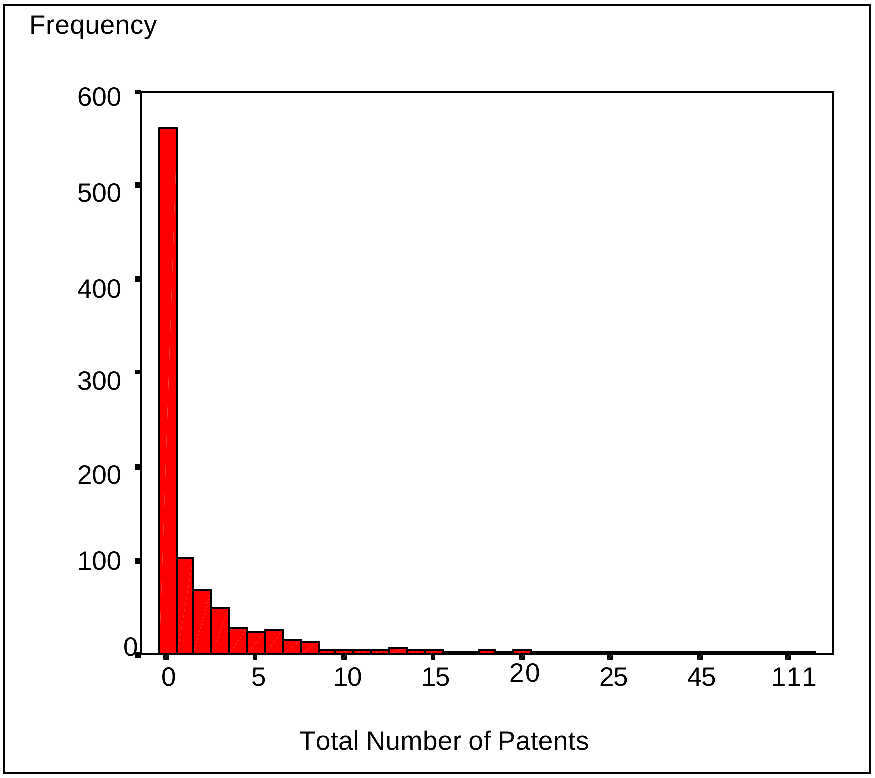

Figure 2. Frequency of Total Number of Patents per Respondent

3.2.6 Jobs Positions and Institutions

The total number of jobs held by any given individual ranged from 0 to 26. This includes all positions and jobs whether or not they were held concurrently. In some cases, for example, a respondent was a full professor, department chair, director of a research center, and an industry consultant simultaneously. This counted as four separate job positions[12]. The mean number of job positions held over the career at the time the data were collected was 6.7 (standard deviation 3.7). The mean number of job institutions was 3.3. As seen in Figure 3, the total number of jobs is skewed right. The average numbers of job positions by job sector were 5.0 academic jobs, 1.2 industry jobs, and .4 government jobs.

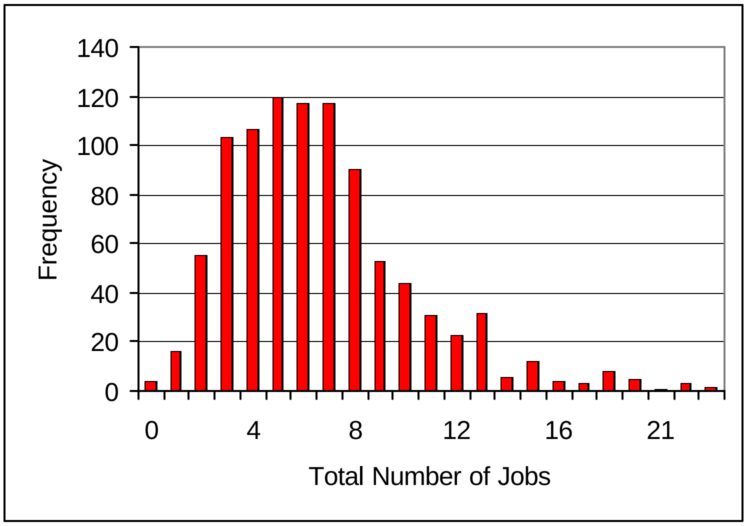

Figure 3. Frequency of Total Number of Jobs Held

3.2.7 Job Transformations[13]

In the RVM dataset as a whole, academic to academic job position transformations (i.e., either a singular or concurrent job position change) accounted for 62.5 percent of all job transformations. Academic to industry job transformations accounted for 4.8 percent of all job transformations. Academic to government transformations represented 2.9 percent of all transformations. For any given researcher in an academic position, the likelihood that the following position will be in academia is .85, that it will be in industry is .07, and that the position will be in government is .04. The remaining .04 represents the frequency of following an academic position within a consulting position (see Table 5).

Industry to industry job transformations accounted for 4.5 percent of all job transformations. Industry to academic job transformations accounted for 8.1 percent of all job transformations. Industry to government transformations represented 1 percent of all transformations. For any given researcher in an industry position, the likelihood that the following position will be in industry is .33, the likelihood that it will be in academia is .59, and the likelihood that it will be in government is .05.

[13] The word “transformations” has been used here rather than “changes” in order to remind the reader that some jobs are held concurrently. The word “changes” suggests that one has left a present position to take on a new position. Transformations means a new job either held singularly or concurrently with other jobs.

The remaining 3 percent represents the frequency of following an academic position within a consulting position (see Table 5).

Government to government job transformations accounted for 1.6 percent of all job transformations. Government to academic job transformations accounted for 3.8 percent of all job transformations. Government to industry transformations represented just 1 percent of all transformations (see Table 5). For any given researcher in a government position, the likelihood that the following position will also be in government is .24, the likelihood it will be in academia is .58, and the likelihood it will be in industry is .15. The remaining 3 percent represents the frequency of following a government position within a consulting position.

3.2.8 Grants

The total number of grants awarded (see Figure 4) to respondents as either principal or co-principal investigator ranged from 0 to 130 with a mean of 17.7 (standard deviation 16.2; skewness of 3.33). Approximately 60 percent had listed no grants on their CVs. For those who had grants, an average of 53 percent of the total were awarded by the federal government compared with 24 percent from industry, and 23 from other sources such as private foundations or state and local entities.

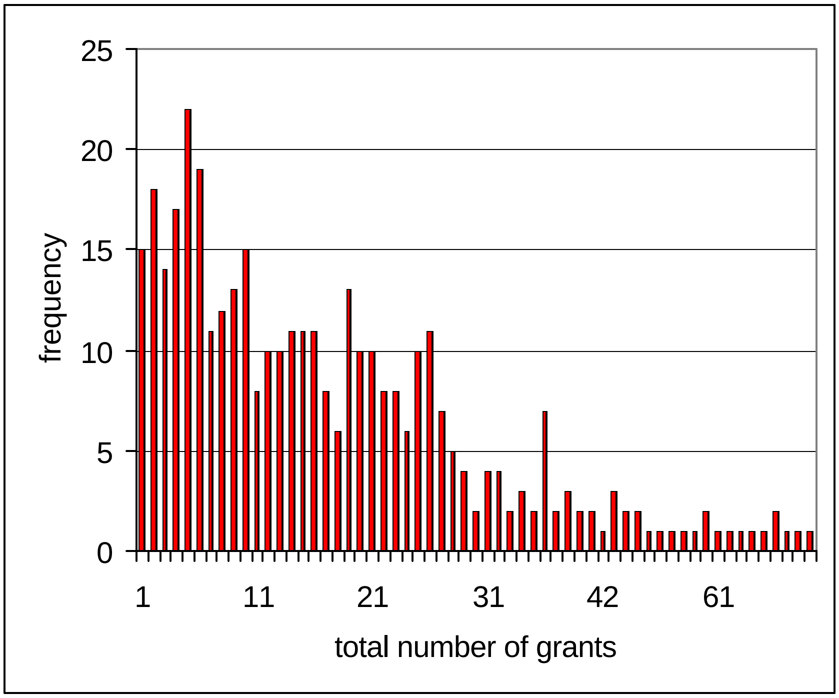

Figure 4. Frequency of Total Number of Grants per Respondent (includes only those with one or more grants)

CHAPTER 4

RESEARCH QUESTIONS AND CONJECTURES

4.1 Research Questions and Analyses

The central research question of this dissertation is "what effects do intrasectoral and intersectoral career patterns have on productivity (as measured in publication and patent counts) over the career life cycle?" The central hypothesis to be tested—the "diversity" hypothesis—is derived from human and social capital theory:

> *Diversity Hypothesis*: "Inter and intrasectoral changes in jobs throughout the career will result in higher research productivity (due to the opportunity provided to build human and social capital)."

The main rival to the diversity hypothesis is the "homogeny"[14] hypothesis:

> *Homogeny Hypothesis*: "Following the 'traditional' career path will yield higher productivity. Scientists and engineers who exhibit a career pattern of relatively uninterrupted job sequences in academia will have higher publication productivity than those who do not. Likewise, those with higher levels of career time in industrial jobs will have higher patent productivity (due to the differences in job incentives between academia, industry, and government)."

[14] A term borrowed from biology meaning "correspondence in form or structure" (Random House College Dictionary, 1984)

Independent of the diversity and homogeny are two additional hypotheses—the "education and training" hypothesis and the "precocity" hypothesis which state:

> *Education and Training Hypothesis*: "Early career experiences through postdoctoral research experiences will result in higher productivity (because they provide educational and human resources (i.e., human capital) building opportunities)."

> *Precocity Hypothesis*: "Scientists and engineers who demonstrate early productivity by publishing before the doctorate will exhibit higher career productivity overall (due either to innate talent, working with highly productive scholars, higher quality graduate training, or other factors)."

The analyses in this dissertation include an examination of (a) basic descriptive statistics, (b) Tobit models of publication and patent productivity, (c) a Poisson model of patent productivity, (d) an analysis of publication and patent productivity "stars" and how they differ from non-stars and from each other, (e) an examination of productivity changes in job transformations to and from industry, and (f) two preliminary Neural Network models.

4.2 Constructs and Conjectures

4.2.1 Diversity of Career Pattern

The effect of the diversity of career pattern on productivity will be measured by variables such as the number of job positions held over the course of a researcher's career as adjusted by career length[15], the proportion of the scholar's career spent in industry and governmental jobs, the first job (industry or government) after the doctorate, and a dummy variable for those who have had

[15] The square of this variable will also be included in the models to detect the possibility of a quadratic relationship between jobs and productivity. That is, the possibility that too many jobs may be detrimental to one's career.

at least one job in all three sectors (named the triple helix). Under the diversity hypothesis it is expected that those with at least some non-zero proportion of the career spent working in industrial and governmental jobs will have higher overall productivity than those with relatively little or no time in industry and government. Likewise, it is expected that the productivity of those who have worked in all three sectors will be higher than those who have not.

4.2.2 Homogeny of the Career Pattern

A variable job "homogeny" was created to index the extent to which a career pattern more or less "conforms" to the norm. All 5,490 job transformations were examined and conditional probabilities were constructed for all possible job transformations. For example, of all respondents who held job type A, what are the relative frequencies with which they proceeded to job type B or C or D, etc.? For each respondent a string of career conditional probabilities was constructed, summed, and divided by the total number of jobs that respondent had held over the career. The result is a variable that measures how far from the norm a respondents career pattern is. The variable has a possible range of 0 to 100. In practice, the range of this variable was .35 to 72.47 with a mean of 14.47 and a standard deviation of 18.10. Individuals with higher values indicate more "typical" the career patterns.

Under the diversity hypothesis it is expected that homogeny will correlate negatively with productivity. The inverse will be true if the homogeny rival hypothesis is true.

4.2.3 Education and Training and Precocity

The education and training and precocity hypotheses are relatively straight forward in terms of variables and measurement. Under the education and training hypothesis those individuals who have had a postdoctoral position[16] are expected to have higher overall career

[16] Initially, graduate assistantships were included in this hypothesis but were deleted from the model when little variance across cases was detected.

productivity than those who have not. Likewise, those who publish the year of the doctorate or earlier are also expected to have higher overall career productivity.

CHAPTER 5

DATA ANALYSIS

In this chapter I analyze the data using quantitative approaches at multiple levels of analysis. A relatively large number of approaches are used, in part, to contribute to the knowledge base about the strengths and weaknesses of these approaches for modeling career data collected from CVs.

In 5.1, I present Tobit models where the dependent variable is the publication or patent rate and the independent variables are those used to test the hypotheses outlined previously. The Tobit approach was chosen to address the censoring problem with career data. In 5.2, a Poisson model (often used with count data) is examined as another approach to assessing patent productivity. Both of these modeling approaches have their benefits and limitations. In 5.3, I examine differences in means on career variables for those classified as publication and patent "stars." These are individuals whose productivity places them in the top ten percent of the distribution in terms of their publication and patent rates. In section 5.4, I examine the productivity effects of making job transformations to and from industry in order to test the diversity hypothesis from another perspective. Finally, in section 5.5, I examine productivity rates by using a non-parametric set of Neural Network models—which are useful when parametric approaches are of questionable value and when data are "noisy."

5.1 Tobit Models of Publication and Patent Productivity

5.1.1 The Basic Tobit Model

Tobit (Tobin, 1958) models are appropriate in cases where data are censored either from "below" (such as observing purchases of durable goods only when respondents having incomes exceeding an arbitrary level are included) or "above" (such as when test scores at the very high end actually do not measure student knowledge appropriately for those for whom the test may be too simple). When samples are censored, ordinary least squares (OLS) regression is not appropriate because they will yield biased and inconsistent estimates—the error term, a function of the explanatory variables, will be non-zero in sum and correlated with the explanatory variables.

In the RVM data set there are several forms of censored data. The primary one is where respondents either do not list the number of publications or list only the most recent. When an item is not specified on the CV, it is difficult or impossible to determine the difference between zero and missing data.

Mathematically, the Tobit model can be expressed as:

$$\text{if } Y_i > T, \qquad Y_i = ß_1 + \mathbf{ß_2 X_{2i}} + u_{2i}$$

$$\text{otherwise,} \qquad = T$$

Where T is a threshold value.

Tobit models use maximum likelihood methods for estimation of model parameters, which yield asymptotically consistent and efficient estimates, given correct model specification (Arminger, 1995). For each non-censored observation, the estimate is simply the height of the density function, representing the probability of getting that particular observation. The probability for each censored observation is the integral above the appropriate density function for threshold level (T) (Kennedy, 1998). Coefficients on the estimated parameters are interpreted much the same as OLS coefficients[17].

The log-likelihood function for the Tobit model is:

$$L = S_{yt>T} \ -½[\ln(2\pi) + \ln(\sigma^2) + (1/\sigma Y_t - X'_t\alpha)^2] + S_{yt=T} \ \ln [1\text{-}F(X'_t\alpha)]$$

[17] The coefficients are interpreted as estimates of the change in the latent dependent variable. That is, the dependent variable as predicted accounting for the censored cases.

Where the first part of the equation ($S_{yt>T}$ -½[ln(2π) + ln(σ^2) + (1/ σY_t – $X'_t\alpha)^2$]) is the standard likelihood function to be maximized, the second part ($S_{yt=T}$ ln [1-F($X'_t\alpha$)]) corresponds to the estimates of the censored cases, and where F is the cumulative normal distribution function.

Several sets of career variables are used to test the hypotheses outlined previously. These include:

1. The Education Set
    - Field of doctorate
    - Did the respondent hold a postdoctoral research position?
2. The Precocity Set
    - How many publications did the respondent have by the year the doctorate was earned?
3. The Job Set
    - Diversity through career start: Did the respondent begin his or her career in industry or government?
    - Triple helix: Did the respondent hold a job in all three sectors (i.e., academia, industry, and government)?
    - Diversity through proportionality: What proportion of the respondents' total jobs years were spent in industry and government jobs (respectively)?
    - Total job institutions: How many organizations was the respondent employed at over his or her career as a function of the number of years of career length?

- Homogeny: Did the respondent exhibit a career pattern that is “typical” or “atypical”?

4. The Research Grant Set

- Resource intensity: How many research grants has the respondent received per career year?

- Resource diversity: What proportion of the grants were awarded by industry or federal sources (respectively)?

5. The Control Set

- Ph.D. age cohorts: Does it matter, even after adjusting for career length in the above delineated variables, when the respondent received his or her doctorate?

- Research center affiliation: Does working in specific research centers make a difference in the productivity of the center’s scientists and engineers?

Thus, the basic statistical models[18] to be estimated are as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} + \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} + \beta_V \mathbf{X_V} + \varepsilon$$

Where,

X_1 = Career homogeny (index of career pattern conformity)
X_2 = Precocity (cumulative number of publications at the doctorate year)
X_3 = Held postdoctoral position (1=Yes, 0=No)
X_4 = Triple helix (had one or more jobs in all three sectors)
X_5 = First job was industry job? (1=Yes, 0=No)
X_6 = First job was government job
X_7 = Years in industry jobs/total job years

[18] The econometrics software *Shazam* was used to fit the Tobit and Poisson models described in this chapter.

X_8 = Years in government jobs/total job years
X_9 = Total number of job institutions/career length
X_{10} = Square of above job variable
X_{11} = Total grants/career length
X_{12} = Industry grants/total grants
X_{13} = Federal grants/total grants
X_{14} = Doctorate granted before 1972? (1=Yes, 0=No)
X_{15} = Doctorate granted 1972-1980
X_{16} = Doctorate granted 1981-1987
X_{17} = Doctorate granted 1988-1993
X_{18} = Doctorate in biological science? (1=Yes, 0=No)
X_{19} = Doctorate in computer science
X_{20} = Doctorate in engineering
X_{21} = Doctorate in physical sciences
$\mathbf{X_V}$ = A vector of dummy variables for each research center

And where,

Y = either the patent rate (number of patents per career year) or the publication rate (number of publications per career year starting the year after the doctorate).

5.1.2 Publication Productivity

Education Variables: Physical and mathematical scientists and engineers were more productive than researchers in other fields. Being a physical or mathematical scientist, is associated[19] with approximately 1.4 additional publications on average per year than researchers from from the reference fields holding all other variables constant. Likewise, engineers had approximately 1.0 more publications per career year than researchers from the reference fileds (see Table 6 for summary statistics discussed in this section).

There was no statistically significant difference in publication productivity between those researchers who have held a postdoctoral research position compared to those who have not, although the coefficient on this term is moderate and negative. For those scientists and engineers in the RVM data set who had a postdoctoral position at some point in their careers, the productivity rate was .31 lower than those who did not, holding all other variables constant.

The coefficient estimate on precocity was statistically significant although the magnitude of its effect was low to moderate. For each paper published by the year of the doctorate, the publication rate after the doctorate increased by .12, holding all else constant.

[19] Unless specified, all results discussed yielded variable coefficients with p-values of .05 or below

Job Variables: The effect of career homogeny on publication rate is low (i.e., the coefficient is .05). This suggests that there may be some mild positive association between career pattern homogeny and publication productivity.

The number of jobs held in different job institutions over the career was included in the model in the form of two variables. The first is the number of job institutions in total divided by the career length as measured in years. The second is the square of the first variable. Under the diversity hypothesis, the number of job positions held was hypothesized to have a positive effect on productivity (due to the human and social capital building opportunities each position may provide). However, the squared term was added to test if having many jobs might actually depress productivity (either due to inability to hold a steady job or due to the time pressures of holding multiple jobs simultaneously). The coefficient on the first order term was negative but not statistically significant[20]. The second order term was also not statistically significant but perhaps not important since the coefficient was low (.05). On the whole this suggests job changes across institutions (of any type) may be associated with lower publication rates, at least for this sample of scientists and engineers.

There seems to be no plausible difference in publication rates, statistically speaking, between scientists and engineers who began their careers in academia versus industry or government. However, for those who began their careers in government, the coefficient estimate on the term was negative and moderate in size (-.54). For those who began their careers in industry, the coefficient estimate on the term was negative but small (-.04). The coefficient for the triple helix career (those who had jobs in all three sectors) was positive (.28) but not plausibly different from zero in a statistical sense.

With regard to job years spent in industry and government, the proportion spent in industry had a moderate negative (-.64) effect on publication productivity, whereas the proportion spent in government had a positive and strong effect (1.15) on publication productivity. However, neither coefficient was statistically significant.

[20] The coefficient estimate was –1.33 (P-value = .08).

Grant Resources: Scientists and engineers who averaged more grants per year appear to have slightly higher publication productivity rates. For each additional grant per year, the number of publications per year increased by .14, although the coefficient estimate was not statistically significant. Although the source of the grant has no statistically significant association with publication rates, for scientists and engineers in the RVM data set, it appears that as the proportion of total grants awarded by industry sources increased by one percentage point, the number of publications per year increased by .17. On the other hand, as the proportion of total grants awarded by the federal government increased by one percent, the average yearly number of publications increases by .66.

Control Variables: The coefficient estimates of all of the Ph.D. cohorts were statistically significant, suggesting that they have higher publication rates than the most recent cohort (Ph.D. earned after 1994), holding all else constant. The normalized coefficient estimates suggests that productivity rates are essentially progressively higher for scientists and engineers in earlier cohorts. Scientists and engineers who earned their doctorate before 1972, for example, published on average .78 more publications per year than researchers in the reference group. This may be due to the likelihood that the dataset may capture a larger proportion of the earlier cohorts' peak productivity years compared to more recent cohorts.

There were six centers[21] where the researchers had higher publication productivity. These are the Biotechnology Process Engineering Center at the Massachusetts Institute of Technology (ERC), the Center for Quantized Electronic Structures at University of California—Santa Barbara (STC), the Center for Dielectric Studies at the Pennsylvania State University (IUCRC), the Center for Ultra-High Speed Integrated Circuits and Systems at University of California, San Diego (IUCRC), the Center for Low Cost Electronic Packaging at the Georgia Institute of Technology (ERC), and the Center for Biological Timing at the University of Virginia (STC). One center had an

[21] It should be noted that although these centers are located at the universities cited, they are often a multi-institutional collaboration representing scientists and engineers at several university campuses.

overall productivity rate that was lower than the others, the Computational Field Simulation at Mississippi State University (ERC).

In summary, among the variables that had statistically significant coefficient estimates, the strongest effects were field differences[22] (with physical and mathematical scientists and engineers having higher publication productivity), cohort differences (where earlier Ph.D. age cohorts exhibited higher productivity), and center affiliation differences (where the scientists and engineers in some centers tended to be more productive than in other centers). Precocity and homogeny both had a weak, positive, relationship with publication rates.

The coefficients on the education variables were not statistically significant, nor were the coefficients on the variables associated with diversity of job patterns or grant patterns. Finally, the overall model was statistically significant (Wald Chi-Square statistic of 179 and P-value of 0.000). The squared correlation between the observed and expected values of the dependent variable was .26.

5.1.3 Patent Productivity

Education Variables: Physical and mathematical scientists, engineers, computer scientists, biologists were all more likely to patent than researchers in other fields (such as social and behavioral scientists and medical scientists) (see Table 8 for summary statistics discussed in this section). The model estimates that physical and mathematical scientists had .53 more patents per year than reference group researchers (holding all else constant), followed by engineers (.53), computer scientists (.46), and biologists (.31). Having had posdoctoral position did not seem to have an effect on patent productivity; the coefficient was small and negative (-.05) but not statistically significant.

The coefficient estimate on precocity was positive but small. For each paper published by the year of the doctorate, the patent rate increased by .02, holding all else constant.

[22] For a summary of publication and patent rates by disaggregated disciplinary fields, see Table 7. In general, chemists, physicists, electrical engineers, and chemical engineers had among the highest publication rates. Among engineers, it appears that civil and mechanical engineers had lower publication rates. Among scientists, it appears that general biologists had lower publication rates.

Job Variables: Unlike the model of publication rate, career homogeny did not have a statistically significant relationship with patent productivity and the coefficient is zero. Like the model of publication rate, the coefficient on the number of jobs held in different institutions over the career was negative and moderate. For the scientists and engineers in the RVM dataset, for each job institution change throughout the career, the patent rate decreases by .18. The squared term had a positive but small coefficient (.05), although the coefficient on neither term was statistically significant.

There seems to be no plausible difference, statistically speaking, among scientists' and engineers' patent rate and the job sector where they began their careers (academia, industry, or government). However, those scientists and engineers in the RVM dataset who began their careers in government had a lower patent rate (-.10) than those who did not, all else constant. For those who began their careers in industry, the coefficient estimate on the term was also negative but small (-.01). The triple helix career (those who had jobs in all three sectors) was positive (.05) but not plausibly different from zero in terms of statistically significance.

With regard to the proportion of job years spent in industry and government, the proportion spent in industry had a strong positive relationship to patent productivity, whereas the proportion spent in government had a negative moderate effect (-.24) on productivity, although this latter coefficient was not statistically significant. As the number of years a researcher spends in industry as a proportion of his or her total career years increases by one percent, the model estimates that the average number of patents per year increases by .83, holding all else constant.

Grant Resources: Unlike the model of publication productivity, scientists and engineers who averaged more grants per year did not have higher patent rates. The coefficient was not statistically or substantively significant. However, the source of grant award does seem to be related to patent productivity. As the number of grants awarded by industry sources (as a proportion of all grants) increases by one percent, the patent rate increases by .41[23]. In comparison, as the proportion from federal sources increases, the patent rate decreases by .06 (although these coefficients were not statistically significant, they nonetheless hold for the scientists and engineers in this sample).

Control Variables: The coefficient estimates of all of the age cohorts were positive, suggesting that they have higher patent rates than the more recent cohort (Ph.D. earned after 1994), holding all else constant. Like publication rate, this may be due to the likelihood that scientists and engineers in earlier Ph.D. cohorts have experienced their prime productivity years whereas those in more recent cohorts may not yet have reached their productivity peak.

There were two centers that a statistically significant higher overall patent rate—The Center for Sensors and Actuators at the University of California-Berkeley (IUCRC) and the Center for the Biotechnology Process Engineering Center at the Massachusetts Institute of Technology (ERC). Researchers at this latter center had on average 1.3 than researchers affiliated with other centers. This effect may be due to the presence of one extreme outlying case, the researcher with the highest number of career patents (141) is affiliated with this center.

There were three centers that exhibited statistically significant lower overall patenting rates. These are the Center for Materials Handling/Logistics Institute at the Georgia Institute of Technology (ERC), the Pacific Earthquake Engineering Research Center at the University of California-Berkeley (ERC), and the Center for Particle Astrophysics also located at the University of California-Berkeley (STC).

In summary, among the variables that had statistically significant coefficient estimates, the strongest effects were field differences[24] (in particular, physical and mathematical scientists and engineers had higher patent rates), cohort differences (where earlier Ph.D. age cohorts exhibited higher productivity), and center affiliation differences (where the scientists and engineers in some centers tended to be more or less productive than in other centers). Unlike publication productivity, patent productivity seems to be positively related to the proportion of total job years spent in

---

[23] P-value = .07

[24] In general, it appears that chemists, electrical engineers, chemical engineers, and physicist had among the highest patent rates. Among engineers, civil engineers seem to have lower patent rates. Among scientists, biologists appear to have lower patent rates. See Table 7.

industry jobs and perhaps to the proportion of all research grants that are awarded by industry sources. Higher overall participation in industry jobs and funding from industry sources tended to have a strong positive association with higher overall patent rate.

Like the model of publication rate, precocity had weak positive effects on patent rates. However, unlike publication rate, the coefficient on career homogeny was not statistically significant. The coefficient on the postdoctoral research position variable was small, negative, and not statistically significant.

Finally, the overall model was statistically significant (Wald Chi-Square statistic of 110 and P-value of 0.000). The squared correlation between the observed and expected values of the dependent variable was .20. See Appendix B for a discussion of model assumptions and diagnostics.

5.2 Poisson Model of Patent Count

5.2.1 The Basic Poisson Model

The problem of censoring may not be as large for patent data as it is for publication data. Patent data were collected from the USPTO for this analysis. The USPTO maintains an online database, which contains data for all patents granted from 1976 to the present. As a result, there may be some left censoring for those scientists and engineers in the RVM dataset who received their Ph.D. before 1976[25]. Of all the researchers in the RVM dataset, 26 percent earned their doctorates prior to 1976, which suggests that their patent counts may be censored from below. However, just 15 percent of these scientists and engineers had any patents at all making them unlikely cases for censoring. In addition, when the data were cleaned the CV was retrieved for those respondents who were granted patents around the time of the USPTO database cut off year of 1976. Previous patents (prior to 1976) were retrieved from the CVs in these cases. Finally,

[25] There are very few cases of patent precocity in the RVM dataset. Scientists and engineers almost never patent before their Ph.D.

Poisson models are thought to be less sensitive to censored samples than ordinary least squares regression. In sum, patent data may not face as many cases of censoring as the publication data.

Poisson models are more appropriate than ordinary least squares regression for estimating models where the dependent variable is measured in counts. First, regression models may produce case estimates below zero (negative values), which is conceptually intractable when working with count data. Second, distributions of count data are often skewed, include many observations of zero count values, and relatively few high value counts. A high number of zero values[26] makes transformation of the dependent variable less likely to produce normally distributed error terms and than when working with continuous data. Unlike the Tobit model, the Poisson model is less sensitive to skewed data and may be more appropriate as a result in the modeling of patent data. In fact, patent data are often used as a case example of the appropriate use of Poisson models (Kennedy, 1998; Arminger, 1995).

The Poisson model assumes that the mean of the dependent variable and its variance are equal, which is advantageous for the modeling of data where the distribution is skewed-right (such as is the case with the patent data). As shown previously, patent data are distributed non-normally with a long right tail. And because the variance is fixed to the mean, as the mean increases because of the presence of highly productive outliers, the variance will also increase.

The basic Poisson model is given by:

$Pr\ (Y_i = y) = e^{-?}?^{y}/y!$ where $? = \exp(\mathbf{X\text{ß}})$

There are two implications of the structural form of the Poisson model. First, it ensures non-negativity in the response variable. And, second it assumes that the errors follow a Poisson distribution (i.e., that error variance increases as a function of its expectation).

---

[26] Approximately 58.5 percent of the scientists and engineers in the RVM dataset had zero patents.

Despite its utility with count data, Poisson models can be used with count data that has been scaled by some variable of interest, such as time periods. In this case, the rate can be interpreted as the expected number of times an event occurs within a specified time period.

Scaling has some advantages in working with Poisson models. With pure count data, models are often “overdispersed.” That is, the variance is often several times larger than the mean[27]. This problem can be ameliorated, to some extent, by creating a ratio[28] variable from the count variable.

In interpreting the coefficient estimates in Poisson models, the sign of the coefficient is and indication of the effect on the expected number of counts. Coefficients with positive signs indicate more counts; negative signs indicate fewer counts, holding all other variables constant. Like logistic regression, to calculate the relative change in rate at which the event occurs for a one-unit change in an independent variable, the coefficient estimate is exponentiated.

5.2.2 The Poisson Model of Patent Rate

As seen in Table 9[29], there were seven variables for which the coefficient estimates were statistically significant—precocity, the number of years working in industry jobs as a proportion of total job years, the number of industry grants as a proportion of total grants, a doctorate granted between 1972 and 1980, a doctorate in the physical and mathematical sciences[30], a doctorate in

---

[27] In these cases a model based on the negative binomial distribution is appropriate. The negative binomial assumes that the variance exceeds the mean (Kennedy, 1998).

[28] It should be noted that count data in ratio form preserves the essence of working with count data in that zero values remain the same. Thus, creating a ratio variable from a count variable is appropriate when using the Poisson model. The mean for patent rate was .143 and the variance was .150.

[29] The model was significant at the .001 level. The R-squared statistic was .10 and the adjusted R-squared was .08. This means that the model does not explain a large fraction of the variation in the dependent variable. The G-squared statistic, a measure of overdispersion, was 337.85, which suggests moderate to low overdispersion.

[30] Because field differences are similar to the results of the Tobit model of patent rate, they will not be discussed here. See Table 9 for the estimates on these variables. In addition, the large numbers of centers dummy variables were not used in this analysis for pragmatic reasons—the

engineering, and a doctorate in computer science. All signs on these variable coefficients are positive; each is associated with higher patent rates.

Among the strongest statistically significant relationships to patent productivity rate is the number of job years the researcher worked in industry as a proportion of total job years. The model estimates that increasing this proportion by one percent, increases the patent rate by a factor of 5.8, holding all else constant. Similarly, increasing the number of industry grants as a proportion of all grants by one percent, increases the patent rate by a factor of 3.7.

There were several variables coefficients that were not statistically significant but are of interest to this analysis. Researchers who had a postdoctoral position were less likely to have higher patent rates than those who did not. In addition, the homogeny index had a small negative coefficient, suggesting no effect or a slight negative effect on publication rate for the scientists and engineers in the RVM dataset, holding all else constant.

Scientists and engineers in the RVM dataset who started their careers in an industry or government job position had lower patent rates than those who began in academia. In fact, their patent rate was only about 91 percent and 85 percent that of their academic counterparts, holding all other variables constant. Those researchers who had at least one job in all three sectors were had patent rates only 82 percent of those who did not. In addition, researchers who spent more of their career years working in government jobs had a lower patent rate than those who did not.

In terms of grant resources, the number of grants that researchers in the RVM dataset were awarded as a proportion of their career length seems to have no effect on their patent productivity. And as the number of grants funded by federal government sources as a proportion of all grants increases by one percent, the patent rate decreases by half.

In summary, the model demonstrates a strong statistically significant relationship between the proportion of total job years spent working in industry jobs and the patent rate, although it appears that starting one's career in industry or government may be detrimental to patent productivity (although these coefficient estimates were not statistically significant).

model failed to run and exceeded the memory partition allowable in the Shazam econometrics software.

The percentage of total grant awards from industry funding sources also was positively associated with patent rate. And it may be possible that scientists and engineers with high proportions of their funding from the federal government have lower patent rates, specific to those in the RVM sample.

There were strong field differences detected for physical scientists, engineers, and computer scientists—all being more likely to patent than their counterparts in other disciplines. Among these physical scientists and engineers had a patent rate more than 5 times that of researchers in other fields.

5.3 Analysis of Publication and Patent Productivity Stars

This section examines the similarities and differences in the careers of the most highly productive scientists and engineers—publication and patent "stars." Publication and patent stars are compared with each other as well as with non-stars. Stars are defined as scientists and engineers who are in the top ten percent of the distribution in terms of their publication and patent rates (i.e., their total number of publications or patents divided by their career length as measured in years). In terms of the scientists and engineers whose career records are included in the RVM dataset, those with a publication rate over 8.2 publications per year are labeled stars for the purpose of this analysis. Likewise, patent stars have .38 patents or more per year of career length. "Combination stars" are those who meet both those criteria. Those who do not qualify in either of these categories are labeled non-stars (see Figure 5). Because of the large number of pair-wise comparisons, I will keep the discussion of findings brief, however, Tables 10-15 show all comparisons.

		Top 10 Percent in Publication Rate?	
		Yes	No
Top 10 Percent in Patent Rate?	Yes	Combination star	Patent Star
	No	Publication star	Non-star

Figure 5. Classification of Stars

5.3.1 Publication Stars Compared with Non-stars[31]

In making a general characterization between publication stars and non-stars in the RVM dataset, it may be said that publication stars had more predoctoral publications, a higher grant rate, a higher degree of career homogeny, and more job positions over the course of their careers than non-stars. They were more likely to have had at least one industry job than non-stars. They were also more likely to be physical scientists and less likely to have had a postdoctoral research position (see Table 10).

In terms of job pattern differences, the career homogeny index for publication stars was 15.3 compared to 14.7 for non-stars, suggesting that publication stars may have more "typical" job patterns, although this difference was not found to be statistically significant. However, publication stars had 1.24 more jobs over the course of their careers on average than did non-stars and more total job years as adjusted for career length than non-stars (3.3 compared with 2.4 for non-stars). This latter difference suggests that publication stars are more likely to have held jobs concurrently. Approximately 24 percent of publication stars had a postdoctoral research position compared with 32 percent of non-stars.

Aside from having approximately 9 more publications per year on average, publication stars also held more patents. They averaged .35 patents per year compared to .05 for non-stars. Finally, in terms of external research grants, publications stars received 1 more grant per year on average per year than non-stars.

[31] Statistical significance of differences in means for all variables can be found in Tables 10-15. A summary of statistically significant differences across all groups can be found in Table 19.

No differences between the two groups were detected in the proportion of job years spent in academia, industry, and government. In addition, publication stars and non-stars were about equally likely to have had at least one job in government. There was also no significant difference in the frequency with which publication stars and non-stars started (i.e., first job) their careers in either industry or government.

5.3.2 Patent Stars Compared with Non-stars

In general, patent stars can be characterized as having had more predoctoral publications, a higher grant rate, a higher proportion of grants from industry, and more industry job positions and job years over the course of their careers than non-stars. They were also more likely to be engineers or physical scientists than non-stars. Patent stars were less likely to have had a postdoctoral research position, had fewer government job positions, proportionately fewer grants from the federal government, and a lower career homogeny rate than non-stars (see Table 11).

In regard to job pattern differences, patent stars were less likely to exhibit patterns of high career homogeny than non-stars. This indicates that patent stars were more likely to make atypical career job transformations than non-stars. The mean of the career homogeny index for patent stars was 12.3 compared with 14.7 on non-stars (14.5 is the overall mean for the RVM population).

Fifty-two percent of patent stars started their careers in an industry position compared to about 30 percent of non-stars, and 72 percent of patent stars had at least one job in industry over the course of their career, compared with 48 percent of non-stars. Patent stars spent about 24 percent of their career years in industry jobs compared to 11 percent for non-stars.

On the other hand, patent stars were less likely to have spent time in government jobs. About 16 percent of patent stars had at least one governmental job compared to 27 percent of non-stars.

In addition to having on average .9 more patents per year than non-stars, patent stars averaged approximately 3 more publication per year than non-stars. Patent stars averaged 1.7 grants per year compared to 1.0 for non-stars. Thirty-three percent of their total career grants were awarded by industry sources compared to 23 percent for non-stars. Conversely, 45 percent of patent stars' grants were from federal sources as compared with 55 percent for non-stars. About 22 percent of patent stars had a postdoctoral research position compared to 32 percent of non-stars.

5.3.3 Publication Stars Compared with Patent Stars

In general, publication stars had higher levels of job homogeny than patent stars, more predoctoral publications, spent a higher proportion of their career years in academic jobs, were more likely to have held at least one job in government, and seem to have more grants per year on average. They also seem to have a higher proportion of the total grant awards from federal sources compared to patent stars. Publication stars spent fewer of their total job years in industry jobs and received fewer of their grants from industry as a proportion of their total number of grants (see Table 12).

In terms of job pattern differences, publication stars tended to exhibit higher career homogeny rates than patent stars. The mean on the career homogeny index for publications stars was 15.3 compared with 12.3 for patent stars. This suggests that publication stars have made job transformations over the career that are more conditionally probable—they are more likely to have followed "typical" career job transformation patterns over time. The mean of the career homogeny index for all scientists and engineers is 14.5, suggesting that publication stars exhibit more career homogeneity than the average. Patent stars, on the other hand, not only have a lower career homogeny index than publication stars but are below the average for all RVM scientists and engineers.

Publication stars were much less likely to have started their career in industry than patent stars. About 29 percent of publication stars' first job was in industry compared to 52 percent of patent stars. Overall, however, this suggests that academic scientists and engineers may more often start their jobs in industry than is commonly thought.

Publications stars were more likely to have had at least one government position during the course of their careers (30 percent had a government job) compared with 16 percent of patent stars. Both groups were about equally likely to have had a postdoctoral research position. Publication stars averaged 7.2 predoctoral publications compared with 4.8 for patent stars.

5.3.4 Combination Stars

In general, it appears that combination stars can be characterized as something of a hybrid of publication and patent stars with just a few differences of note (see Tables 13-15). Combination stars (16.8) appear to have a higher career homogeny index than publication stars (15.3), who have a higher index than non-stars (14.7), who have a higher index than patent stars (12.3). Combination stars (8.4) also appear to have a higher number of predoctoral publications than publication stars (7.2), patent stars (4.8), and non-stars (3.2). Combination stars are less likely to have held government jobs than publication stars and non-stars but less likely than patent stars to have held an industry job.

5.4 Productivity Changes in Job Transformations to and from Industry

To test the assertion that intersectoral changes in jobs affects overall productivity, all of the transformations from academic jobs to industrial jobs and vice versa were identified and the mean number of publications was calculated for the five years before and after the transformation. These means were then summed and averaged over all job transformations made by all scientists and engineers between academia to industry. As seen in Table 16, the mean number of publications for the five-year period preceding a job transformation *from industry to academia* (for all scientists and engineers who made this transition) was 1.5 publications per year. For the five-year period after the transformational move to academia, the mean was 2.6 publications—a 1.1 increase in the mean number (and a statistically significant difference at the .05 level). Thus, the average productivity of scientists and engineers increased after a job transformation from industry to academia.

A similar analysis was performed for all job transformations made by the scientists and engineers *from academia to industry*. For the five-year period preceding the move, the mean number of publications overall was 1.8; for the five-year time interval following the transformational move to industry, the mean was 2.6—an increase of .8 publications per year (also a statistically significant difference. Thus, the mean number of publications for the five-year period of time following a move from academia to industry also increased.

These results may suggest that making a job transformation has important effects on productivity. However, an alternative interpretation is that making a job change is correlated when productivity rates are naturally increasing. For example, it may be the case that researchers in the early stages of their careers are more likely to make job transformations than those in later stages and that this is precisely the same time period when researchers' productivity is on the rise.

5.5 Neural Network Analysis

In this section, I present two "preliminary" Neural Network models of publication and patent productivity. This work is deemed preliminary because it is an initial investigation into these methods and the subject of future work and refinement[32]. I include it in this dissertation as a contribution to the potentially useful methodological approaches to analyze career patterns and productivity.

5.5.1 The Basic Feedforward Backpropagation Neural Network Model

Neural Network models view the independent variables as "inputs" and dependent variables as "outputs." Between the inputs and outputs are one or more hidden layers of processing entities known as "neurons" that receive input weights and generate transfer weights. By comparing these weights with the output target values, through massive computer processing cycles, the weights are adjusted until they come close to their target output values. There is no underlying assumption of the distribution of data, nor are variables "held constant" during weighting

[32] Because the preliminary character of this work, discussion of the data analysis will be kept intentionally brief.

iterations or cycles—the goal is to minimize global error through a process known as backpropagation. As weights are changed through iterative cycles so too are the modeled interrelationship among inputs and outputs.

In Neural Network models, outputs depend on patterns of inputs and are thus robust toward outlying or extreme cases. The processing entity (discussed below) usually employs a "squashing" function. In the case of feed-forward backpropagation models, such as used in this research, the squashing function is most often sigmoidal in shape where extreme values are confined within the limits of zero and one. Neural Network models are known to be robust toward coding errors, and missing or noisy data (Garson, 1998, p. 9).

Because there are no assumptions about the underlying form of distribution of data (i.e., a non-parametric method), Neural Network models are particularly useful in modeling data that are nonlinear and where assumptions of normality cannot be met (Garson, 1998). As argued previously, Neural Networks may be a useful way to examine and model publication and patent rates due to the non-normal, skewed distributions of these data.

A Neural Network is based on multiple neurons, also called processing entities or nodes, arrayed in one or more layers, which pass data among each other and adjust a set of weights according to a specified mathematical (squashing) function. The neurons are interconnected and store information based on the weights assigned to the interconnections. These weights are adjusted to refine the predictive or classificatory precision of the model. Links between nodes in the input layer, the hidden layer(s), and the output layer cannot be reduced to a simple equation such as in regression modeling, since inputs and hidden nodes are linearly or nonlinearly interdependent.

The input weights are measures of the relative importance of the connection between the input neurons and the neurons in the hidden layer (these will be displayed below). Weights are initialized to random values in the training process but become more meaningful as they are adjusted in the learning process (i.e., the cyclical adjustment of weights). Positive weights denote the connection is excitatory; they are negative if the connection is inhibitory.

A summation function adds the weights and computes a net input to each neuron. For the target (output) neuron, the net input is the sum of the path weights from all of the input neurons times the outputs of these neurons:

$$\text{Net input}_j = b_j + Sw_{ji}O_i, \text{ where } b_j \text{ is a bias weight.}$$

The summed weights are fed forward to the next layer of neurons by a transfer function (also called activation function), which invokes a mathematical expression[33] to convert the summed weighted inputs into a transfer weight (also called activation value). If this transfer weight meets or exceeds a threshold level (i.e. the learning rule set by the researcher), the neuron passes a signal forward in the form of an output weight.

Backpropagation is the process by which error is reduced through the use of an error gradient, which updates the weights. The new weight for the connection between an input neuron and a hidden layer neuron is previous weight plus the learning rate times the transfer weight times the error term for neuron (the difference between the output value and its target value). This new weight can also be adjusted by changing the momentum rate (which is the degree to which information in the previous cycle is retained). This process proceeds until the adjustments are smaller than some criterion established by the researcher[34].

---

[33] The transfer function usually introduces nonlinearity into the neural network model, such as the sigmoidal logistic function ($1/e^{(-\text{net input})}$).

[34] In the case of this research, the criterion was an average overall error (the difference between the target and output values) below two percent. Approximately one-third of the data was used in training; the rest used in validation. Through optimization, the learning rate was set at 1.0 and momentum at .8.

5.5.2 Neural Network Model of Publication Rate[35]

Among the influential inputs in modeling publication rate (as a single output) are: (1) precocity, (2) number of years in government jobs as a proportion of total job years, (3) number of years in industry jobs as a proportion of total job years, and (4) career homogeny (see Figure 6 and Table 17). Each of these is described below.

[35] In the case of this model, the criterion to terminate learning was set to an average overall error (the difference between the target and output values) below two percent, which was reached after 1,680 learning cycles. Approximately one-third of the data was used in training; the rest used in validation. The learning rate was optimized at 1.0 and momentum at .80.

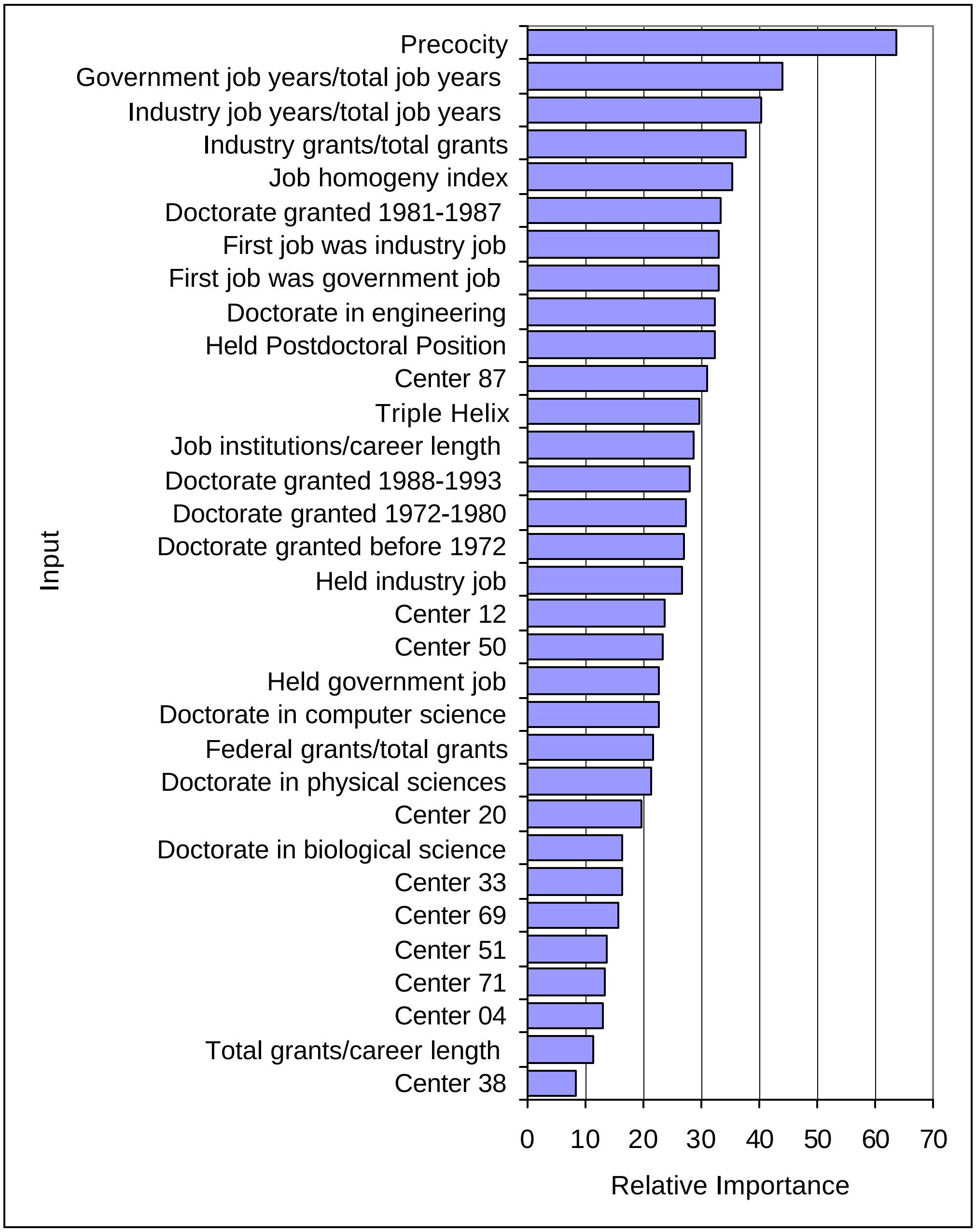

Figure 6. Neural Network Model: Relative Influence of Inputs in Modeling Publication Rate

First, the relationship between precocity and publication rate appears to be non-linear but strongly positive. The estimated effect of precocity on the (after-doctorate) publication rate seems to be moderate through up to approximately 15 predoctoral publications but then increases at a strong rate, especially between 15 and 20 publications (where the slope is steepest) (see Figure 7).

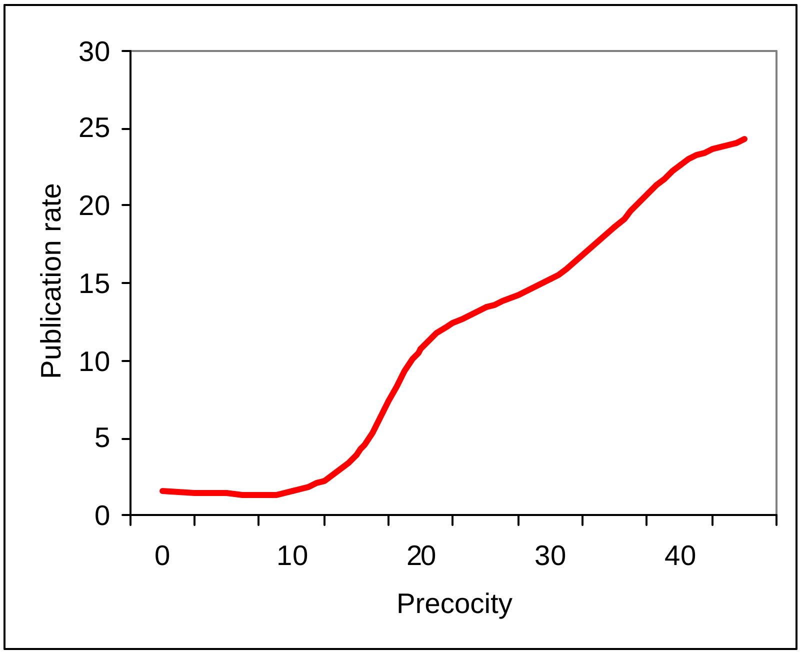

Figure 7. Estimated Relationship Between Precocity and Publication Rate

Second, the model estimates a negative relationship between the number of years a researcher worked in government jobs as a proportion of total job years. The negative relationship seems to bottom out at around 80 percent but is sharply negative until that point (see Figure 8).

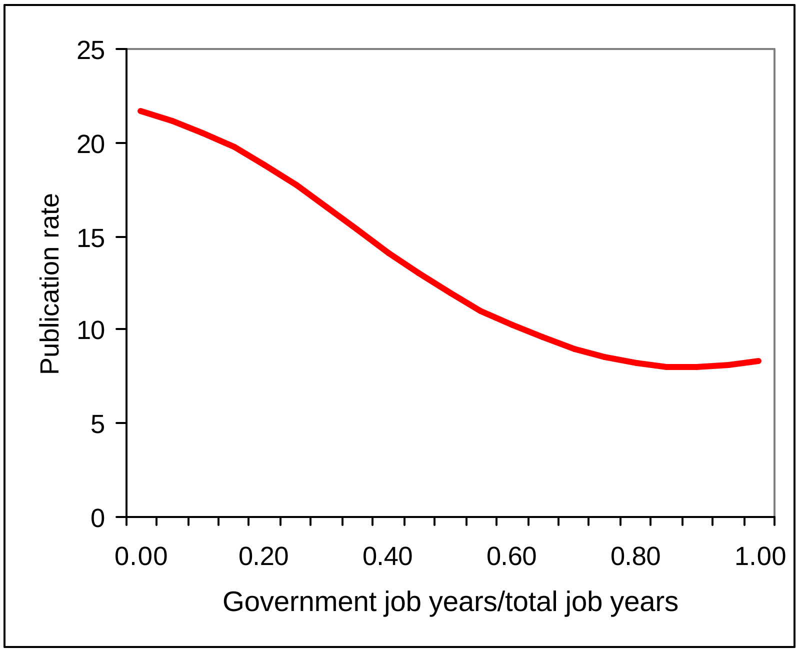

Figure 8. Estimated Relationship Between the Number of Years in Government Jobs as a Proportion of Career Length

Third, the model estimates a negative relationship between the number of career years spent in industry as a proportion of total career years and the publication rate. The effect is large for those who spent between zero and approximately 50 percent of their job years in industry and then begins to decrease (see Figure 9).

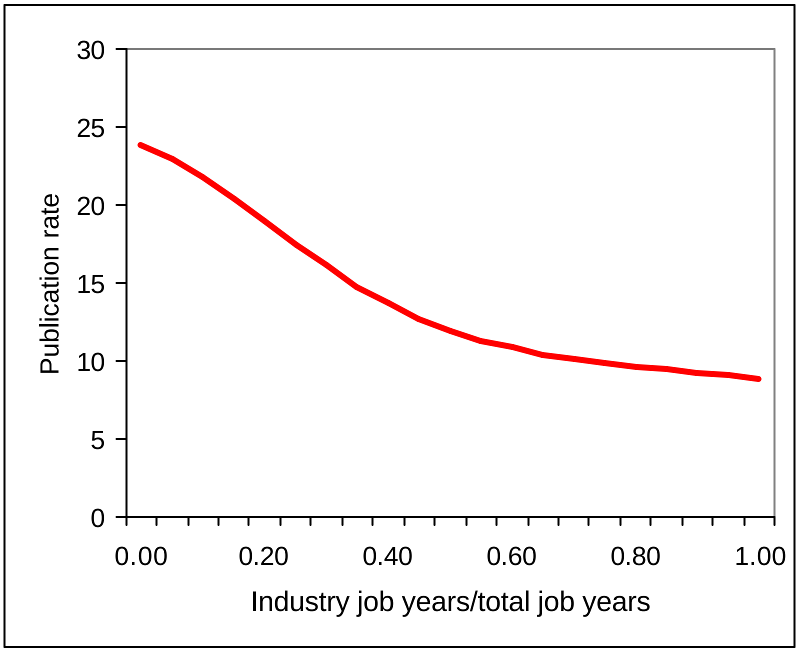

Figure 9. Estimated Relationship Between Industry Years as a Proportion of All Job Years and Publication Rate

Fourth, career homogeny is estimated to be positively related to publication rate, but highly non-linear and "S" shaped in form. For those with career homogeny indexes up to approximately 30, the relationship is essentially flat (the overall mean for homogeny is 14.5). Between approximately 30 and 50, the relationship with publication rate is strongly positive and then flattens out again (see Figure 10). So those researchers with extremely high levels of career homogeny are associated with extremely high publication rates.

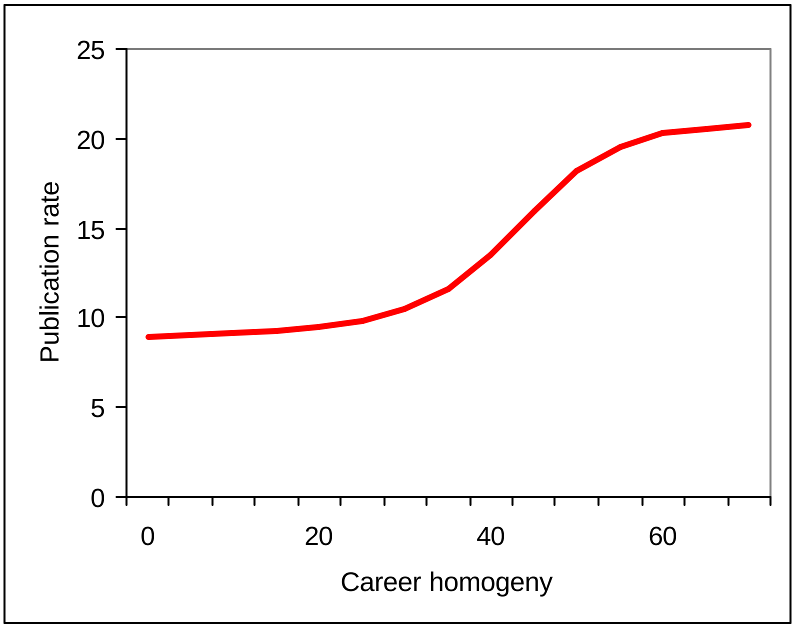

Figure 10. Estimated Relationship Between Career Homogeny and Publication Rate

Finally, the model estimates that those who have had postdoctoral research positions have much lower publication productivity rates (approximately 10 publications per year) than those who did not have postdoctoral positions (20 per year), holding no other variables constant, however.

5.5.4 Neural Network Model of Patent Rate[36]

In brief, the patent model indicates that (1) affiliation with Biotechnology Process Engineering Center at the Massachusetts Institute of Technology, (2) career homogeny, (3) a doctorate awarded between 1988 and 1993, (3) an earned doctorate in engineering, (4) a postdoctoral research position, and (5) held one or more industry job positions are among the most influential inputs (see Figure 11). Each of these relationships is addressed below.

[36] In the case of this model, the criterion to terminate learning was set to an average overall error (the difference between the target and output values) below two percent, which was reached after 880 learning cycles. Approximately one-third of the data was used in training; the rest used in validation. The learning rate was optimized at 1.0 and momentum at .80.

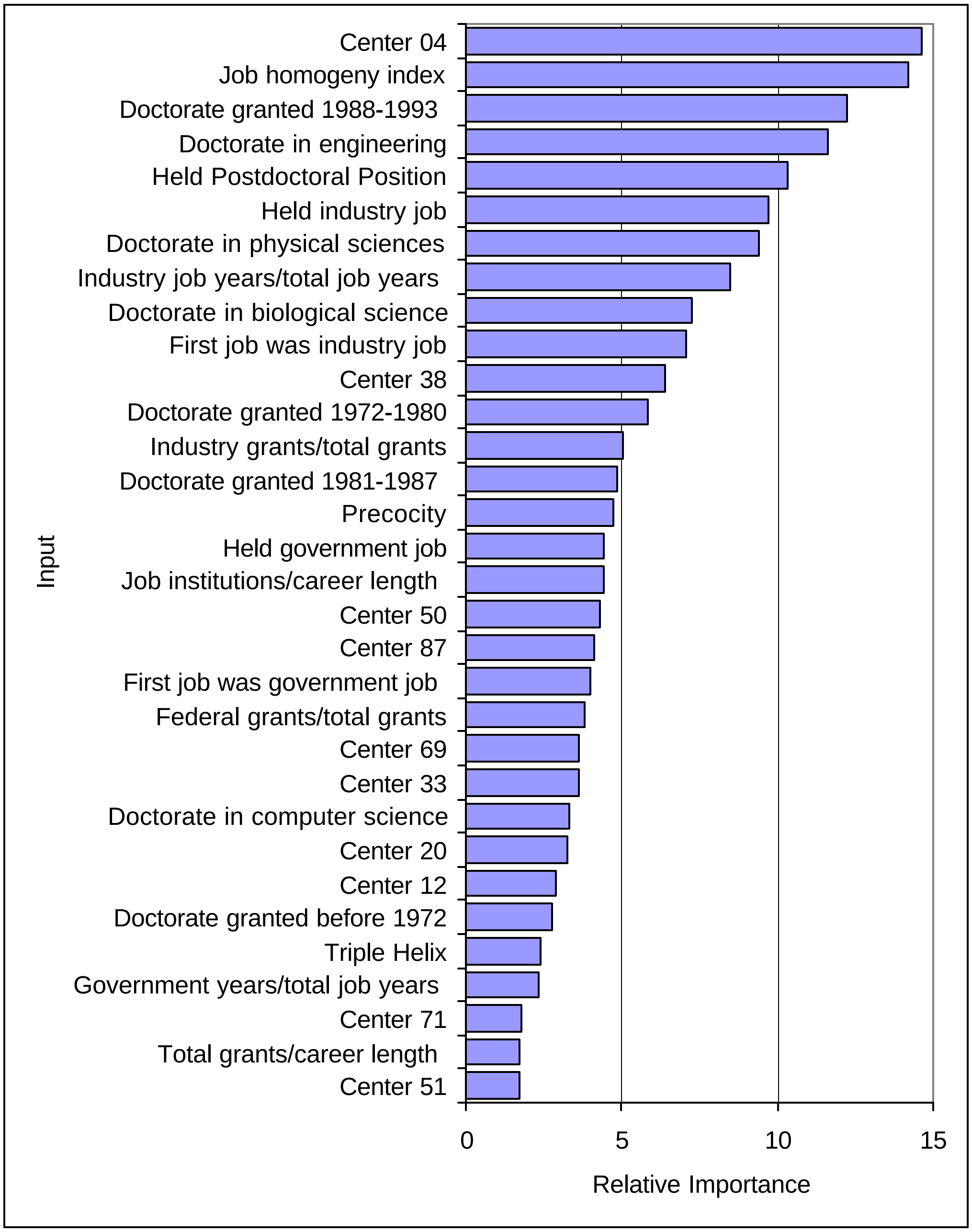

Figure 11. Neural Network Model: Relative Influence of Inputs in Modeling Patent Rate

First, the relationship between career homogeny and patent productivity is estimated to be positive but in a highly curvilinear fashion. That is, for researchers who had a career homogeny index up to 35 or 40 (where the mean is approximately 14.5) the model estimates that the average number of patents per year hovers close to zero. However, researchers that exhibited extremely high career homogeny index values are estimated to also have extremely high patent rates (see Figure 12).

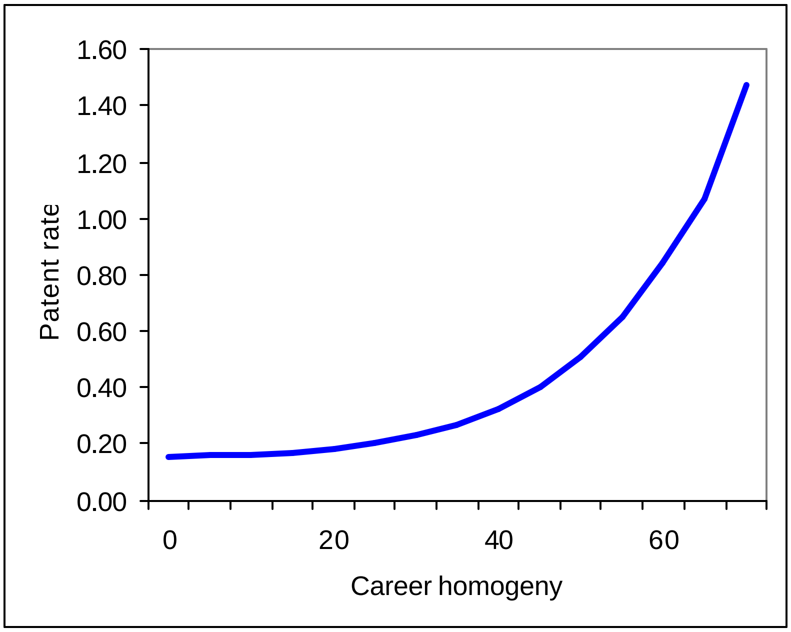

Figure 12. Estimated Relationship Between Career Homogeny and Patent Rate

Second, the model estimates that those at the ERC located at MIT averaged about 1.4 patents per year compared to .10 per year for those researchers affiliated with other centers. Third, engineers averaged about .56 patents per year compared to .24 for those in other fields (see Table 15), holding no variables constant.

Fourth, the model estimates that those who have held postdoctoral positions averaged about .14 patents per year compared to .77 for those who did not.

Finally, the model estimates that those who have had at least one industry job have higher patent rates than those who did not. For those with one or more industry jobs, the model predicts a patent rate .86 of as compared to .16 for those who had no industry jobs. However, there is a positive, curvilinear relationship between the number of years spent working in industry jobs as a proportion of total job years. The effect is moderate until about 50 to 60 percent and then rises sharply (see Figure 13).

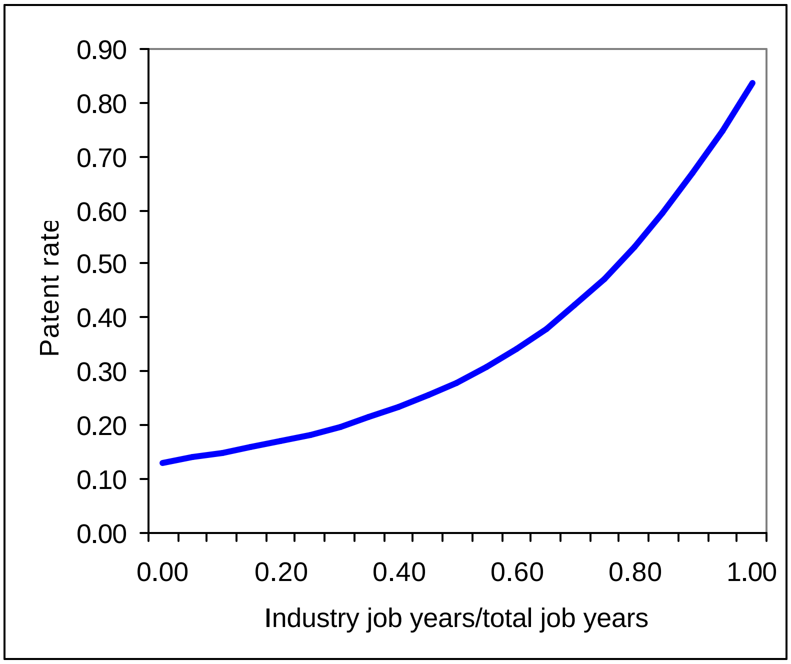

Figure 13. Estimated Relationship Between Years in Industry Jobs as a Proportion of Total Job Years

Finally, the model estimates that the relationship between precocity and patent rate is also highly nonlinear and positive. This relationship does not seem to be substantially different for researchers with relatively high numbers of predoctoral publications (e.g., 5-30) than it is for researchers who with few or no predoctoral publications. In contrast, for those with roughly 30 or more predoctoral publications, the patent rate begins to increase sharply (see Figure 14).

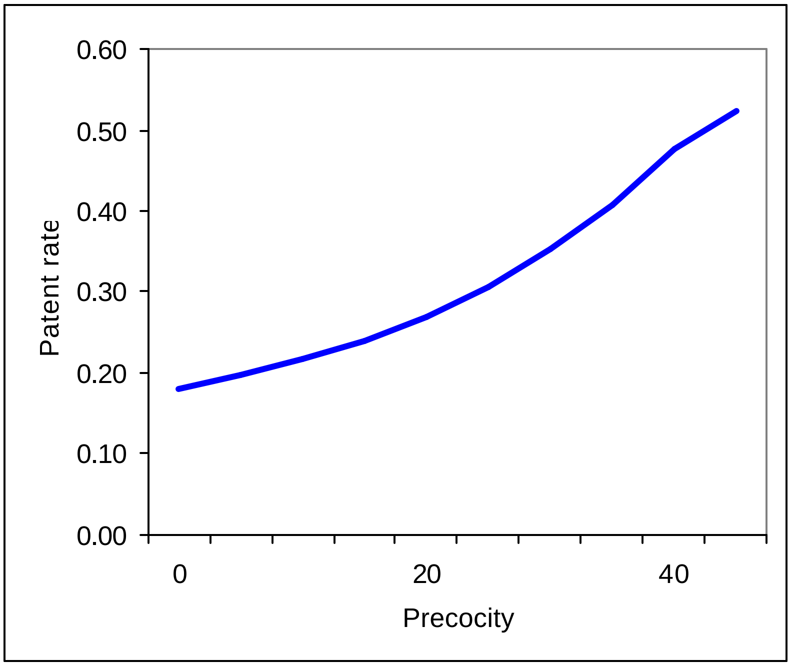

Figure 14. Estimated Relationship Between Precocity and Patent Rate

5.5.5 Summary of Neural Network Analysis

Neural Network analysis has been employed to demonstrate its usefulness in modeling interactive, nonlinear, and nonparametric data such as productivity data. The curvilinear relationship between variables such as career homogeny and precocity and productivity are consistent with the small coefficient estimates found in the Tobit and Poisson models—the Neural Network model has helped to identify why. In general, career homogeny was among the most influential inputs in each of the models, but in each case (and particularly for patent productivity) only extremely high values of homogeny were associated with high productivity rates. In both models, having had a postdoctoral position was negatively related to productivity rates. Finally, precocity was positively associated with productivity in both models, but seemed to be less influential in the modeling of patent productivity. In both models, however, only relatively high numbers of predoctoral publications were associated with high levels of productivity.

CHAPTER 6

DISCUSSION OF FINDINGS

In this chapter, I interpret the results of the analyses reported in the previous chapter and discuss their implications for the hypotheses as posed in Chapter 5. Because several statistical models have been used, with different model assumptions, the results differ and sometimes conflict across models. These will be noted and discussed. Section 6.1 focuses on publication productivity while section 6.2 addresses patent productivity. Section 6.3 summarizes the differences and similarities in findings for publication and patent productivity. Finally, section 6.4 interprets the evidence for and against the hypotheses as originally posed.

6.1 Publication Productivity

Publication productivity was modeled in a Tobit model and a Neural Network model. In addition, I examined differences in means on a host of variables between publication stars and non-stars and analyzed the relationship of job transformations to and from industry with publication productivity.

6.1.1 Education and Human Resource Variables

Field of Doctorate: From the Tobit model (see Table 7 for coefficient estimates and Table 16 for a summary of the effects of variables across all models) being an engineer or a physical or mathematical scientist was related to higher publication productivity. This is consistent with the findings from the Neural Network models (see Table 17) that estimate that these two fields are associated with higher publication rates than other fields. A higher proportion of physical and mathematical scientists were publication stars than was true to patent stars and non-stars (although they were about equally likely to be engineers) (See Table 19 for a summary of differences between publication stars, patent stars, and non-stars).

Postdoctoral Position: Across all of the models, the effect of having had a postdoctoral research position appears to be negative. The coefficient on this variable in the Tobit model was negative and moderate although not statistically significant; the Neural Network model estimated a strong negative effect. In addition, publication stars were slightly less likely to have had a postdoctoral position than non-stars (24 percent versus 32 percent).

This finding suggests that postdoctoral positions may be ineffective in terms of helping researchers to establish productive research careers over the long term. It is possible that they have shorter term impacts on productivity or that they have other positive effects such as providing financial access to scientific and technical education and careers. It is also possible that these are crude proxy measures of what ultimately may matter—the quality of the experience, working with a productive mentor and the culture of the work environment as a place of learning and human resources development. Postdoctoral positions may be highly variant in terms of these qualities, which may explain this counterintuitive result. Yet both of these findings are policy relevant due to the large public expenditures on supporting such positions. As a result, this finding deserves further exploration and may require a more thorough study involving the use of theories from the science of learning and better empirical measures.

Publication Precocity: Precocity (or predoctoral publications) showed only a slightly positive effect on future publication productivity in the Tobit model. For every additional predoctoral publication, future publications increased by an average of .12 per year. This is consistent with the Neural Network model finding that precocity seems to have no substantial association with future publication productivity except for high values of precocity (i.e., when a researcher had 15 or more predoctoral publications). Moreover, publication stars averaged approximately 7.2 predoctoral publications compared with 3.2 for non-stars. This suggests two things. First, predoctoral publication is not as rare as may be commonly believed and, second, extremely high productivity in the predoctoral years may indicate higher overall career productivity, but otherwise may not have much effect.

Perhaps precocity is a better measure of retention in the profession, quality of future work, likelihood of making tenure, or short-term productivity. But consistent with Reskin's (1977) finding, there is little evidence to suggest that precocity is associated with long-term (or career) productivity rates. This may also be consistent Long's and McGinnis' (1985) finding that the principal predoctoral effect on long-term productivity is tutelage with a productive scholar.

6.1.2 Job and Career Variables

Starting Position: It appears that starting one's career in government is associated with lower overall publication productivity across the career cycle, at least for those researchers in the RVM dataset. All models indicate that, in general, researchers who began their careers in government had lower publication rates, although none of these estimates was statistically significant. Starting one's career in industry seems also to be negatively related to future publication productivity. The Tobit model showed a slight negative relationship and the neural model estimated substantially lower publication rates for those who started their careers in industry jobs.

However, although not substantively different, approximately 29 percent of publication stars' and 30 percent of non-stars' first job was in industry. This suggests a higher rate of early career mobility between industry and academia than may be commonly believed.[37] It may be the case that these proportions are higher than in the overall population of scientists and engineers due to the industrially oriented mission of the centers in this study and the relative overrepresentation of engineers and physical scientists in this data set. Yet the finding is noteworthy, particularly in understanding the "academic" career. The extant literature on academic careers has taken little note of this and has examined academic job changes, which may have solidified assumptions of little mobility between sectors.

[37] At least this is the case for this sample of researchers.

Proportion of Career in Industry and Government: It appears that the proportion of a researcher's career spent in industry is slightly negatively related to publication productivity and the proportion spent in government jobs is slightly positively related (although the coefficients estimates in the Tobit model were not statistically significant). The Tobit model estimates that as the proportion of the career spent in industry increases by one percent, the estimated publication rate decreases by .64. Conversely, as the proportion in government jobs increases by one percent, the estimated publication rate increases by .15. These findings are not consistent with the Neural Network model where the model estimated a negative relationship between years in industry and government and publication productivity.

Job Transformations To and From Industry: In examining the effect of job transformations to and from industry on productivity, it appears that productivity increases both after a move to industry as well as after a move from industry. For the five years before a job position move *from* industry, researchers averaged 1.5 publications per year compared to 2.6 for the five-year period following the move. Similarly, in job transformations *to* industry, researchers averaged 1.8 publications per year as compared to 2.6 for the five-year period following the move. There are two possible interpretations of this finding. One possibility is that such job transformations (or changing or adding a job position) truly do boost productivity because of access to new forms or human and social capital as originally hypothesized (or for other reasons). The second possible explanation is that job transformations tend to occur in the period of the career when productivity is on the rise in general, regardless of job location.

However, even if the latter interpretation is correct or partially correct, it is still notable that publication rates increase despite commonly observed industrial disincentives (or the lack of incentive) to publish.

Triple Helix: Across all analyses there appears to be mixed evidence of the relationship between having had at least one job in all three sectors (academia, industry, and government) and publication productivity. I argued that these three job sectors may provide diverse and complementary forms of human and social capital, which would motivate productivity. However, the coefficient on this variable in the Tobit model was small and not statistically significant. The neural model yielded fairly high positive estimates. The stars analysis shows that about an equal proportion (16 percent versus 12 percent) of publication stars and non-stars have had jobs in all three sectors. Most of the evidence here suggests no effect on productivity thus making the original argument difficult to support.

Career Homogeny: The Tobit model estimates no substantive relationship between career homogeny and publication rate. In the Neural Network model, homogeny has a strong positive relationship with publication rate, but only for high values of homogeny. Publication stars had slightly higher career homogeny indexes than non-stars, which is consistent with the Tobit result.

So overall it appears that following a more typical (or more highly conditionally probable) set of job transformations does affects career publication productivi ty, but only in the case of extremely homogenous career patterns. This does not seem to be due to low variance in the data—the mean on the career homogeny index was approximately 14.5 and the standard deviation was 8. The career homogeny variable does not take into account the sequence of jobs per se, it is a chain of conditional probabilities of job transformations divided by the number of jobs for each researcher. It may be the case that sequence is more important than pure probability, although they are not unrelated in that odd sequences would yield low values on the career homogeny index. Perhaps this finding will be explored more thoroughly in future research using Hidden Markov Models (as discussed in Chapter 8) that predict the likelihood of a certain outcome based on a chain of event statuses.

Holding Jobs in Multiple Job Institutions: There appears to be a negative relationship between the number of job institutions of the researcher and publication productivity. The Tobit model estimates a fairly strong relationship. For each additional job institution, holding all else constant, the career publication rate decreases by 1.3. Yet, this coefficient estimate was not found to be statistically significant. The Neural Network model is consistent with the Tobit finding. The analysis of stars suggests that publication stars averaged about the same number (3.6) of job institutions as non-stars (3.3).

6.1.3 Grant Resources

Grant Rate: The Tobit model estimates a slight positive relationship between the number of grants a researcher has received as a function of his or her career length. As the number of grants per year increase by one, the estimated average publications per year increases by .14. This is consistent with the analysis of publication stars. Publication stars averaged about two grants per year compared with one for non-stars. This is also consistent with the Neural Network model.

Proportion of Grants from Industry and Government: Under the diversity hypothesis it was assumed that research support from both industry and government motivates scholars in different ways due the differences in mission, funding requirements, and incentives across the two sectors. For example, industry may be more likely to support work that seeks answers to use-driven questions where the findings have some potential for short-term application, whereas federal support may be directed toward longer-term, curiosity-driven problems or mission oriented problems. The homogeny hypothesis suggests that federal support (being historically the largest source of academic research support) may yield higher publication productivity.

Again, the findings conflict across statistical analyses of the source of grant support. The Tobit model suggests that that there is a slight positive association between the proportion of total grant support that was awarded by industry sources and publication productivity and a moderately positive association between the proportion awarded by federal sources and publication productivity (although neither finding is statistically significant). The analyses of publication stars and non-stars show no meaningful differences.

The Neural Network analysis suggests that the publication rate has perhaps a bimodal relationship with industry grant support. Researchers with a relatively small proportion and researchers with a relatively high proportion of their support from industry are estimated to have higher publication rates than researchers with between approximately 30 and 70 percent of their research grant support from industry (see Figure 15).

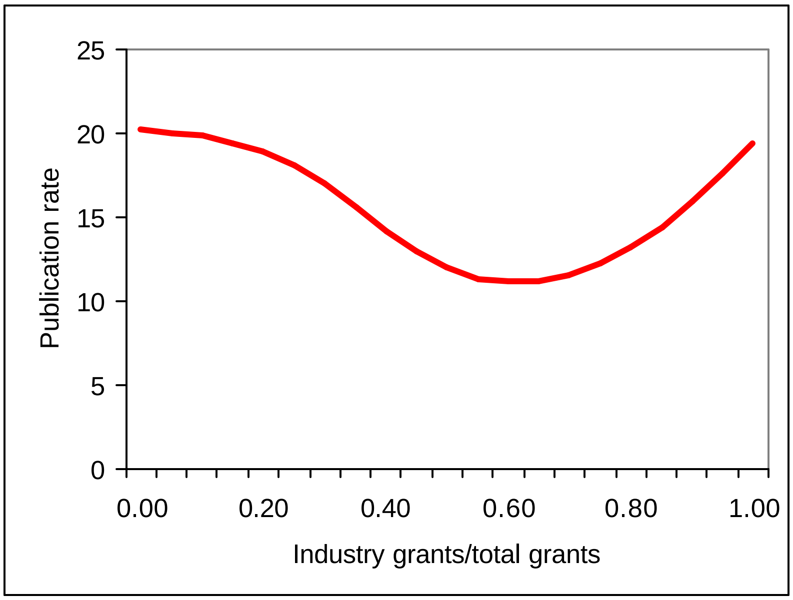

Figure 15. Estimated Relationship Between Industry Grants as a Proportion of Total Grants and Patent Rate

The model estimates a positive, approximately linear relationship between the proportion of support from the federal government and publication rate, although the effect appears to be moderate.

Thus overall it appears that grant support does relate positively to publication productivity but it does not appear to have a strong effect. There appears to be conflicting evidence about the relationship between the source of support and publication productivity—the stars analysis suggests no difference, the Tobit analysis in general support of the homogeny hypothesis, and the Neural Network analyses perhaps suggest some benefit from mixed sector support.

6.2 Patent Productivity

Patent productivity was modeled in a Tobit model, a Poisson model, and a Neural Network model. In addition, I examined differences in means between patent stars and non-stars across the variables of interest to this analysis.

6.2.1 Education and Human Resource Variables

Field of Doctorate: From the Tobit model (see Table 8 for coefficient estimates and Table 18 for a summary of the effects of variables across all models) being a physical or mathematical scientist, an engineer, a computer scientist, or a biological scientist were all related to higher patent productivity than the reference group, although the largest coefficient estimates were for physical and mathematical scientists and engineers. This is generally consistent with the estimates from the Poisson model (see Table 9) and findings from the Neural Network models (see Table 17) where physical and mathematical scientists and engineers were estimated to have higher patent rates and computer scientists and engineers to have lower patent rates. A higher proportion of engineers and physical and mathematical scientists were patent stars than was true of non-stars (although the difference between patent stars and non-stars was not statistically significant for physical and mathematical scientists). A substantially lower proportion of patent stars (2 percent) were biological scientists compared with non-stars (12 percent) (see Table 11 and Table 18 for a summary of differences between patent stars and the other groups).

Postdoctoral Positions: The evidence regarding the effect of having had a postdoctoral research position is consistent across models. The coefficient on this variable in the Tobit and Poisson models was negative and moderate in size (but not statistically significant). This was also true of the estimates from the Neural Network models and is consistent with the comparison of patent stars and non-stars. Just 22 percent of patent stars had a postdoctoral research experience as compared with 32 percent of non-stars (a statistically significant difference). There may be several explanations for this finding. First, it may be the case that postdoctoral positions are often located in the academic sector where patent productivity is not rewarded like publication productivity. Second, although less likely, it may be the case that postdoctoral associates are viewed as research "labor" and are not credited with break through discoveries because their work is dominating by so called "mopping up" duties. Third, postdoctoral positions may be more common in fields where patenting is less common. Finally, postdoctoral positions may have a form of career "holding pattern" effects where top graduates are sought after for assistant professorships and others are left waiting in the career queue.

Publication Precocity: Precocity (or predoctoral publications) showed only a slightly positive effect on patent productivity in the Tobit and Poisson models. In the Tobit estimate, for every additional predoctoral publication, patents increased by an average of .03 per year; the Poisson model estimates a similar result. This is also consistent with the Neural Network models—precocity seems to have a weak positive association with patent productivity except for extremely high values of precocity (i.e., when a researcher had 20 or 30 or more predoctoral publications). Moreover, patent stars averaged approximately 7 predoctoral publications compared with 10 for publication stars, and 3 for non-stars.

6.2.2 Job and Career Variables

Starting Position: It appears that starting one's career in government is associated with lower overall patent productivity across the career cycle, at least for those researchers in the RVM dataset. All models showed that researchers who began their careers in government had lower patent rates, although none of these estimates was statistically significant. Among the stars—just 9 percent of patent stars began their careers in a government job compared to approximately 15 percent for publication stars and non-stars.

There is some evidence that starting one's career in industry may be detrimental to long-term patent productivity. The Tobit model showed a zero relationship, while the Poisson model yielded a small to moderate negative estimate, although neither estimate was statistically significant. The neural models showed slightly higher estimates of patent rate for industry starters.

Yet, in contrast, the stars analysis shows that patent stars were much more likely to have started their careers in industry. Approximately 52 percent of patent stars' first job was in industry compared to about 29 percent for publication stars and 30 percent for non-stars (all of these differences were statistically significant). These findings taken together suggest two observations of interest. First, at least in the early career years, there is substantial job mobility between industry and academia. Second, starting one's career in industry may be related to productivity spuriously through the disciplinary field of the researcher.

Proportion of Career in Industry and Government: Across all of the analyses, the proportion of one's career spent working in industrial jobs appears to have an enormously positive effect on patent productivity. This is the strongest single predictor of patent productivity in the Tobit and Poisson models. The Tobit model, for example, estimates that increasing the proportion of one's career by one percentage point increases the patent rate by .83, holding all else constant. The stars analysis is consistent with this finding. Patent stars spent approximately 24 percent of their career years in industrial jobs compared to approximately 9 and 11 percent for publication stars and non-stars, respectively.

Conversely, a higher proportion of years in government jobs is associated with a lower patent rate, although these coefficient estimates were not statistically significant. These findings are consistent with the Neural Network models. The stars analysis demonstrates that patent stars averaged approximately the same percentage of their total job years in government jobs as publication stars and non-stars.

Triple Helix: The effect of having had a least one job in each of the three sectors (academia, industry, and government) differs across models. The Tobit model estimates no substantive relationship, while the Neural Network model and the Poisson model the estimate is moderate and negative, suggesting that the triple helix career is negatively related to patent productivity (although the Tobit and Poisson coefficient estimates were not statistically significant).

This is consistent with the analysis of patent stars who were less likely to have had a job in all three job sectors (10 percent) compared to non-stars (15 percent).

Career Homogeny: There is some conflicting evidence about the relationship between career homogeny and patent productivity. Both the Tobit and Poisson models yielded near zero coefficient estimates (the latter was not statistically significant). However, the neural models yielded estimates of a positive relationship between homogeny and patent productivity but only for extremely high levels of career homogeny. In contrast, the analysis of stars suggests that patent stars exhibited lower career homogeny indexes (12.3) than non-stars (14.7) and still lower than publications stars (15.3) (all of these differences with patent stars were statistically significant).

Holding Jobs in Multiple Job Institutions: There appears to be a negative relationship between the number of job institutions of the researcher and patent productivity. The Tobit model estimates a fairly strong relationship. For each additional job institution, holding all else constant, the career publication rate decreases by .18. Yet, this coefficient estimate was not found to be statistically significant. This is inconsistent with the Poisson model which estimates that with each additional job institution, the patent rate increases by 30 percent. The Neural Network model is consistent with the Tobit finding. The analysis of stars suggests that patent stars averaged about the same number (3.2) of job institutions as non-stars (3.3).

6.2.3 Grant Resources

Grant Rate: The Tobit and Poisson and Neural Network models estimate no relationship between the number of grants a researcher has received as a function of his or her career length and patent rate. This is not consistent with the analysis of stars, which showed that patents stars averaged approximately 1.7 grants per year compared to 1.0 for non-stars, and with the Neural Network models.

Proportion of Grants from Industry and Government: The Tobit and Poisson models suggests that that there is a large positive association between the proportion of total grant support that was awarded by industry sources and patent productivity and a moderate negative association between the proportion awarded by federal sources and patent productivity (although the only the Poisson estimate on industry funding is statistically significant). The analysis of patent stars is consistent with this finding. About 33 percent of patent stars' grants were funded by industry compared with about 23 percent for non-stars. Similarly, about 45 percent of patent star's support was awarded by federal sources as compared to 54 percent for non-stars. Yet the Neural Network analyses suggests that the patent rate remains relatively consistent for researchers with relatively low proportions of industry support but increases at an increasing rate. The Neural Network model estimates slightly increasing patent rates as the proportion of support from the federal government increases.

Thus overall it appears that numbers of grants per year may be unrelated to patent productivity. However, it seems that the weight of the evidence across models suggests that the source of support does matter—where industrial support is associated with higher overall patent rates than federal support.

6.3 Similarities and Differences in Publication and Patent Productivity Analyses

6.3.1 Similarities in Publication and Patent Productivity Findings

Taken as a whole, the analyses of publication and patent productivity reveal some consistent findings. First, physical and mathematical scientists and engineers have higher productivity rates than researchers in other fields.

Second, publication precocity also seems to have little effect except in cases of extremely high levels of precocity. All of these education and human resource variables taken together suggest long-term, or career, productivity rates are not affected by these variables as measured. Again, it is possible that effects of education and early career variables on productivity could have been detected had more precise variables been available. But in order to do proper justice to the topic, this would need to be the subject of an entire study. Alternatively, it may be that education and early career experiences have important career effects in areas other than productivity.

Third, there is a consistent finding across models that having had a postdoctoral research position is negatively related to career productivity rates.

Forth, across all models there was a consistent finding that researchers who started their careers in government had lower overall productivity. And there seems to be no effect of having had a job in each of the three job sectors examined—academia, industry, and government.

Fifth, there is some evidence that researchers who made relatively high numbers of institutional job changes have lower productivity than those with fewer.

6.3.2 Differences in Publication and Patent Productivity Findings

The major differences in publication and patent productivity involve job and grant related variables. First, there is some evidence that patent productivi ty is positively related to closer connections to industry and negatively related to closer connections with government. Higher proportions of the career spent in industrial settings and higher proportions of research funding from industry sources seem to be related to higher patent productivity. In contrast, higher proportions of the career spent in government jobs and a higher proportion of federal support seem to be negatively related to patent productivity or have no substantial effect.

In contrast, a higher proportion of the career spent in industry seems to be negatively related to publication productivity, while more career years in government jobs seems to be positively related to publication productivity.

Second, a higher number of grants per year is related to higher publication productivity but not to patent productivity.

Third, the source of the support (industry or federal) is unrelated to publication productivity, while a higher proportional level of industry support is positively related to patent productivity.

Fourth, at least within the community of scholars deemed productivity stars in this dissertation, it seems that publication stars exhibited higher career homogeny than patent stars who exhibited lower career homogeny than non-stars[38]. Combination stars had the highest career homogeny indexes. Thus, it seems that following typical career trajectories is related more to publication productivity than to patent productivity. Moreover, publication stars are more likely to hold more job positions per career year than non-stars and possibly patent stars. Finally, patent stars tended to start their careers in industrial positions at much higher rates than publication stars and non-stars and less often in government positions.

6.4 Evidence in Support or Opposition to Hypotheses

6.4.1 Precocity Hypothesis

This hypothesis states: "Scientists and engineers who demonstrate early productivity by publishing before the doctorate will exhibit higher career productivity overall (due either to innate talent, working with highly productive scholars, higher quality graduate training, or other factors)." There is evidence to support this hypothesis, although the relationship between precocity and future productivity appears to be weak and perhaps influenced by several cases of extreme precocity outliers.

The Tobit models estimate that for each additional predoctoral publication, the publication and patent rate increase by .12 and .02, respectively, holding all else constant. To put this in context, the mean publication rate was approximately 4 (publications per year) and the mean patent rate was .14 for all scientists and engineers in the RVM dataset. Proportionate to their means then, the effect of precocity on patent rate was actually estimated to be higher. The Poisson model provides consistent results.

The analysis of stars demonstrates that publication stars averaged approximately 10.2 predoctoral publications, compared with 7.3 for patent stars and 3.1 for non-stars[39]. However, the

---

[38] Although in the analyses of the complete sample career homogeny seems to have a slight positive effect on both forms of productivity overall, mostly due to those with extremely high levels of homogeny.

Neural Network analyses suggest that the effect was essentially zero for those with relatively low to moderately high numbers of predoctoral publications. The relationship was strongly positive for those with more than approximately 20 or more predoctoral publications. In sum, this hypothesis is supported given the above caveats.

6.4.2 Education and Training Hypothesis

This hypothesis states: "Early career experiences through postdoctoral research experiences will result in higher productivity (because they provide educational and human resources (i.e., human capital) building opportunities)." Across all models—Tobit, Poisson, Neural Network, and the stars analysis—there is evidence to support the notion that postdoctoral research positions may inhibit longer-term productivity. The Tobit and Poisson model coefficient estimates were negative and moderate in size. And, publication stars and patent stars were less likely to have held postdoctoral appointments than non-stars. The Neural Network models estimate that publication and patent rates are lower for those who have had a postdoctoral position than they are for those who have not. In sum, this hypothesis is not supported.

6.4.3 Homogeny Hypothesis

This was the main rival hypothesis to the diversity hypothesis discussed in the next section. It states: "Following the 'traditional' career path will yield higher productivity. Scientists and engineers who exhibit a career pattern of relatively uninterrupted job sequences in academia will have higher publication productivity than those who do not. Likewise, those with higher levels of career time in industrial jobs will have higher patent productivity (due to the differences in job incentives between academia, industry, and government)."

There is evidence for and against this hypothesis. I will discuss the evidence for this hypothesis in this section and will discuss the evidence against this hypothesis in the context of the

[39] All of the coefficient estimates on the Tobit and Poisson models were statistically significant at the .05 level or below. The stars analysis shows that the difference in means between publication stars and non-stars was statistically significant.

diversity hypothesis (in the next section) since they are essentially complements. First, the Tobit model of publication rate estimates that the coefficient of the career homogeny index[40] is positive and statistically significant, although small in size (.05). This means that as the homogeny index increases one point, holding all else constant, the estimated publication rate increases by .05. In contrast, for patent rate, the Tobit and Poisson models estimate a coefficient of zero (not statistically significant). The Neural Network models suggest steadily a relatively flat relationship between homogeny and publication and patent rates until the homogeny index reaches approximately 40. So based on the career homogeny variable, there seems to be some evidence for a weak positive relationship between homogeny and productivity except for researchers who have had a career of highly probably job sequences, and then the relationship is strongly positive.

Second, all three models estimate a negative relationship between starting one's career in industry or government and productivity (both for publication and patent rate). Third, for publication rate, the Tobit model estimates a negative relationship with the proportion of career years worked in industry jobs, which suggests a more traditional academic career path results in higher publication rates[41]. Likewise, all models showed a strong positive relationship between proportion of career spent in industry jobs and patent rates, which would be expected under the homogeny hypothesis. Finally, publication productivity is associated with more grants in general, although the source of the grant (federal versus industry) does not seem to matter. Again, this is expected in environments where publication productivity is prized. Overall, there is sufficient evidence to suggest that the homogeny hypothesis is plausibly supported by the data.

Third, all models estimate a negative relationship between the number of job institutions and productivity rates. That is, the more job institution changes a researcher seems to make is related to lower overall productivity.

[40] The range on this variable was from .4 to 72.5

[41] It should be noted that the coefficient on the proportion of the career spent working in government was positive and of moderate size although not statistically significant.

6.4.4 Diversity Hypothesis

The diversity hypothesis—the central hypothesis of this dissertation—states: "Inter and intrasectoral changes in jobs throughout the career will result in higher research productivity (due to the opportunity provided to build human and social capital)." The primary evidence for this hypothesis (and thus against the homogeny hypothesis) is from the descriptive statistics and comparisons of means from the analysis of stars. First, about 29 percent of publication stars began their career in industry and averaged about 10 percent of their job years in industry and 5 percent in government. This about the same proportion as was true of non-stars but suggests some intersectoral career diversity is common. In addition, publication stars were more likely (56 percent) to have had at least one industry job than non-stars (48 percent), although this difference is not statistically significant.

Second, higher patent rates were associated with careers proportionally higher in industry jobs and job years in industry. But even among patent stars the proportion was still relatively modest—24 percent of career years in industry. Thus both publication and patent productivity are associated with some degree of career diversity.

Third, the patent rate was also associated with proportionally higher levels of funding from industry. However, even among patent stars, the proportion of grants awarded by industry sources was roughly 33 percent compared to 45 percent of the awards from federal sources. Incidentally, of publication stars's grants support approximately 28 percent came from industry, compared with 40 percent of combination stars and 23 percent of non-stars.

Fourth, there is evidence from the stars analysis to suggest that homogeny may be related more highly with publication productivity than patent productivity. Career homogeny was highest among publication stars and combination stars. Patent stars had lower homogeny than both these groups and lower than non-stars as well. So there may be reason to conclude that career homogeny is a more important factor in publication productivity than in patent productivity. Thus, the diversity hypothesis may be more relevant in explaining patent productivity as compared to publication productivity.

Finally, researchers in the RVM data set averaged 17 percent of their total jobs in the industrial sector and 12 percent of their career years in industry jobs. Approximately 51 percent had one or more jobs in industry; 26 percent had one or more jobs in government. Nearly half began their careers in non-academic jobs—for 33 percent the first job was in industry and for 15 percent the first job was government. An average of 24 percent of the grants awarded to researchers in the RVM dataset came from industry. These latter statistics suggest that there is perhaps substantially more intersectoral linkages than may be commonly perceived (at least for this dataset of researchers working in academic research centers).

Although the career homogeny hypothesis may be better supported by the data analysis two important findings are drawn from examining the data through the lens of the diversity hypothesis. First, there seems to be substantial intersectoral diversity among researchers in the RVM dataset, perhaps more than would typically be expected. And, second, the diversity hypothesis is more plausible in explaining patent productivity than it is in explaining publication productivity.

CHAPTER 7

LIMITATIONS AND BENEFITS OF USING THE CV FOR CAREER RESEARCH

Few studies have employed CVs as data sources about trends in job patterns and productivity. Typically, when CVs are used in career studies, they are used as a supplemental source of information that serves to fill in gaps left from other research modalities (Long et al., 1993; Gomez-Mejia and Balkin, 1992). Even when CVs[42] are used as the primary or only data source (Bonzi, 1992), their advantages or disadvantages are rarely discussed. The utility of CVs as a data source is dependent upon their quality and completeness. This chapter is designed to address some practical issues when using the CV the approach.

Section 7.1 focuses on the general limitations of using the CV as a data source for career research studies. Section 7.2 focuses on specific issues related to "problem" variables set commonly included on the CV, while section 7.3 addresses the benefits of this unusual method.

7.1 General Issues in Using the CV as a Source of Career Data

7.1.1 Semi-standardized Formatting

Although CVs tend to conform in academia toward a semi-standardized format, there are many differences in the presentation and ordering of information, making data entry difficult, as coders must flip back and forth among the pages of the CV in search of the next variable sets that are demanded of the coding protocol. Generally speaking, name and institutional affiliations are placed first, followed education and degree information, job degree histories, and lists of publications. After that, information tends to come in any order including, lists of affiliations with

---

[42] Methodologically, CVs have been found to closely match information from other secondary sources such as the American Psychological Association's directory. Nevertheless, these other secondary sources have been shown to undercount the number of published journal articles as compared to CVs (Heinsler and Rosenfeld, 1987).

professional societies, grant information, patent information, information on student advisees, courses taught, honors and distinctions, service work to the field, and service work within the university.

Even within these broad categories there are a number of formatting challenges, in some CVs publications are ordered by date (with no specification of the type of publication), some are ordered by type in a number of ways (e.g., refereed journal publications, edited volumes, monographs, peer reviewed conference papers, conference papers, books, book chapters, monographs, etc.), and some ordered chronologically within type.

Another formatting challenge is that some CVs are ordered chronologically (from earliest job or publication, for example) to most recent but some are in reverse chronological order. Within the topic of time sequencing, some CVs order jobs by the year in which they began (straight sequential ordering), while others nest this information within job category types (see hypothetical examples 1 and 2 for a comparison).

Example 1. Straight sequential ordering

Professional Positions:

1970-present	Professor, MIT
1970	Visiting Professor, Max Plank Institute
1966-1979	Director of Materials Research Laboratory
1966	Consultant, General Motors Corporation
1965-1970	Associate Professor, MIT
1964-1972	Adjunct Professor, Tokyo University, Japan
1960-1965	Assistant Professor, MIT

Example 2. Sequential ordering within type

Academic Professional Positions:

1970-present	Professor, MIT
1970	Visiting Professor, Max Plank Institute
1965-1970	Associate Professor, MIT
1960-1966	Assistant Professor, MIT

Other Positions Held:

1966-1979	Director of Materials Research Laboratory
1966	Consultant, General Motors Corporation
1964-1972	Adjunct Professor, Tokyo University, Japan

Both career examples contain identical data although the formatting of the information is different. This is not an insurmountable coding problem but it makes for higher item complexity in the coding task.

7.1.2 Missing Information

Missing information poses one of the greatest challenges to the use of the CV as a data source for career studies. The primary difficulty is in ascertaining if the information is missing as opposed to zero in value. So if, for example, a scientist or engineer has listed no research grants, does this mean that the information is missing or that the person has truly not been awarded a grant? This is the principal problem that makes models such as Tobit analyses and event history analyses well suited to career research as derived from CV data.

Usually, one can develop an informed “hunch” about whether a particular CV contains missing information versus data values of zero by examining the career age of the respondent. If a respondent is a full professor but has no publications, it is likely that this is a case of missing data. However, there is no reliable way to detect the difference between missing and zero data on the CV. One way to address this is to classify the CV as “suspect” when counterintuitive missing information is identified. Then one can choose to include or not include these cases depending on the objective of the research and analysis planned via the use of CV data.

7.1.3 CVs of Varying Length

Related to the issue of missing data is the problem of variance in length of the career and thus the length of the CV. Depending on the career age and accomplishments of the scientist or engineer it is not uncommon for the CV to be as short as one or two pages or as long as 200 pages or more. As a result the number of variables can range from as few as 10 to 3,000 or more. Most standard statistical analysis packages will view these empty variables as missing data when they are, in reality, null data. For the scientist with a career history of only 10 variables there may actually be no missing data at all although modeling software may treat the 26th job type variable as missing for this case.

7.1.4 Truncated CVs

Closely related to missing information is the problem of "truncated" CVs. Most commonly when truncation it is a problem is truncation from below. In these cases, the CV may list only recent information, such as "publications since 1990," or may list only significant information, such as "most significant publications." However, truncated CVs come in other forms such as the narrative from (e.g., biosketch) where information must be extracted from text. In addition, often academic scientists and engineers will have a short and a long form of their CV and it is not always simple to detect if the CV in hand is the more complete one.

Truncation from above comes in two forms. The most common is that the CV has not been kept up to date, so recent information is unavailable. The other form is where the CV does not itemize publications, for example, stating only "over 40 publications."

7.2 Specific Problems with Certain Variables Included on CVs.

7.2.1 Educational Data

Generally, educational data are easy to extract from the CV. However, the most frequent problems include: missing year in which degree was earned and missing field of study. The field in which the bachelor's degree was earned was missing on approximately 10 percent of all CVs in the RVM dataset and the year in which it was earned was missing on 7 percent. Disciplinary field of doctorate was missing on 12 percent of the CVs and the year in which the degree was earned was missing in 5 percent.

7.2.2 Job Data

The most common problems with job data include difficulty in identifying and classifying the job type and missing information in beginning and end dates of job durations. As for difficulty in discerning job type position, there are two problems—difficulty in determining the nature of the position and difficulty in classifying and coding them as a result.

Most academic jobs are relatively easy to identify particularly the typical ones such as assistant, associate, and full professor, or department chair, or dean. Yet others are more difficult, some graduate assistantship positions are difficult to classify as such in that they may be disguised under titles like research associate. This is a specific example of the case of "embellishment" of the CV for job marketing purposes. Or it may be the case that the CV does not distinguish between research assistantships and teaching assistantships, which in theory should be classified differently because of the different skill building tasks they address.

Visiting positions are often difficult to classify correctly. Some are taken as research sabbaticals and are more similar to full professor positions, for example, while others are "soft money" positions where the respondent has no permanent job and may be more closely related to adjunct faculty positions or lecturer positions. Research scientists in some institutions may be tenured research "faculty" and in other institutions they may be also in soft money positions. Finally, there are a number of academic administrative positions that are difficult to distinguish from each other in a human capital or theoretical sense depending on how the title is presented, such as the array of assistant and associate directors, deans, vice presidents, and provosts.

In industry, problems include difficulty in classifying and coding various junior to senior research scientist and engineering positions across firms and within categories. Some job positions in industry distinguish between research scientist or engineer, senior research scientists or engineer, project or program director, and group leader. From a classification point of view, those differences may seem meaningful, but what should be the correct coding classifications, which distinctions are meaningful, and where does one draw those distinctions?

Another difficulty comes in the form of distinguishing between research and administrative positions. Frequently, titles such as group leader, project chief, research director, vice president for research (versus vice president for other matters) are difficult to classify appropriately. Should these be coded as administrative positions or research positions? At what point in the continuum between pure researcher and pure administrator should the line be drawn?

The major problem with consulting jobs is two-fold; they are either underspecified or overspecified. I suspect that some researchers simply do not list consulting work on their academic CVs. When the CV does list consulting jobs it is often difficult to know how to code or "weight" these variables. Does consulting at a company for one day, one week, two months, or three years make a difference to a researcher's career? At the extremes, the answer is likely to be yes in the social and human capital formation opportunities that consulting work carries. The difficulty is in determining what is worth coding when myriad consultancies are listed and coding time is finite.

7.2.3 Publication Data

Publication data from CV records exhibit several data quality problems. First, it can be difficult to determine if a publication has been peer reviewed. Although many scholarly journals are peer reviewed, some are not. In addition, some CVs do not make this information clear and it can be difficult to identify a publication as a peer reviewed journal article compared to some other form of publication, especially when the task includes coding hundreds of CVs, some of which contain hundreds of publications, in just about every field of science and engineering.

Second, there are differences in the quality of the journals so it may not be the case that publications should be counted as equal. Publication counts can be thought of as a relatively crude measure of scientific contribution. Perhaps a better measure of publication quality (or usefulness of the scientific contribution) is to examine citation counts. However, the task of collecting and matching this data can be onerous.

Third, there are a host of disciplinary field differences that interact with publication data problems. In some fields, it is standard practice to have the lead author listed first, in others the ordering is alphabetical, and in still others the senior member of the team is listed first or last. In some fields refereed conference papers such as those embodied in variance proceedings (e.g., IEEE) are highly rewarded, whereas in other fields conference papers are regarded as relatively unimportant and are not peer reviewed. At times it is difficult to determine which is the case.

7.2.4 Patent Data

The most common problem with patent data is that they are frequently not included in the CV. Just 21 percent of the CVs in the initial coding (i.e., pre-data cleaning) contained information about patents, which was found to be an underestimate. By using the USPTO database, and through careful matching of records against the CV job institution at the time the patent was applied for, it was found that 41 percent of the scientists and engineers in the RVM data set had been granted one or more patents.

7.2.5 Grant Data

Like patent data, it is suspected that CVs contain a lot of missing data for research grant awards. Unfortunately, unlike patent data, there is no simple way to test this. Approximately 40 percent of the scientists and engineers in the RVM dataset listed one or more (external) research grant on their CV.

The most challenging problem with grant data is that it is perhaps the least standardized element of the CV in terms of what information is presented and in what format. Common problems include missing duration of the award, unclear identification (including the use of field specific acronyms) of the sponsoring organization, missing information about the dollar amount of the award (including some cases where the total amount is listed, others where direct costs are listed, and others where the “share” applicable to the respondent on a multi-investigator award is identified), and unclear identification of the respondent as principal or co-principal investigator.

7.3 Benefits of the CV Approach

The above litany of difficulties suggest that much more labor and time must be spent in coding, checking, cleaning, and reordering the data. Sometimes this requires reexamining primary sources such as the CV and any relevant website the respondent may maintain. Despite these challenges, the CV remains by far the most complete longitudinal record of scientists' and engineers' careers.

There are several major benefits in working with CV data. First, the method is relatively unobtrusive. It is unlikely that respondents would be willing to provide the level of detail contained in the CV through a survey questionnaire or interview. The enormous time required of the respondent would make that impossible, whereas with data collected via the CV, the time required for the respondent may be just a minute or two to read the email request and to reply with an attached CV file.

Second, the wealth of longitudinal career analyses that one could perform using CV data is enormous including studies such as this one, evaluative and comparative studies of research centers, potential social network studies on smaller samples, as well as studies of gender differences and differences in national origin. The researcher can always return to the primary source (i.e., the CV) to collect supplemental data as necessary—for example, to code and analyze co-authorship patterns and to collect information that would permit further bibliometric analyses. Often with survey research methods this is impossible.

Third, while CVs can sometimes include missing data, they are often kept as a log of sorts and updated with some frequency. This diminishes well known "telescoping" problems (incorrectly identifying the time period and sequence of past events) in survey research and interviewing (Sudman and Bradburn, 1982).

Finally, CV data can be viewed as complementary to more traditional forms of social science data collection and can serve to enhance methods such as site-visiting and face-to-face interviews. Having the CV in hand prior to the interview could result in less respondent burden or may allow the interviewer to focus on richer topics. In the RVM Program, site visits preceded the collection of CVs, which were followed by and informed the construction of a survey questionnaire designed to address topics—such as opinion items, perceptual and behavioral items, and value expressions—not captured by the CV or site visit.

CHAPTER 8

LIMITATIONS OF STUDY AND FUTURE RESEARCH DIRECTIONS

Aside from those delineated in Chapter 8 on the method of CV collection, coding, and analysis, there a number of limitations of this study. In addition, there are a number of avenues for productive future research. Both are discussed in this chapter. In sections 8.1 through 8.4 the limitations of this research are addressed. In section 8.5 possible areas for future research are discussed.

8.1 Sample Limitation

As noted in Chapter 4, the sample framework for this study was non-random. Individuals were chosen because of their affiliation with research centers funded by the agencies that supported the RVM Program. As a result, the findings pertain only to this population of scientists and engineers. In addition, despite best efforts, the response rate was low (36.5 percent) and there may be problems with non-response bias as a result. For example, it might be the case that the most productive scholars receive many email messages that they cannot or choose not to respond because of other time-use priorities.

In contrast, the reported response rate was based on a sample framework that included undergraduate and graduate students as well as some administrative support personnel (neither the former nor the latter were used in this study). These individuals may not yet have a CV or may have believed that they were incorrectly identified as intended respondents due to the language used in the recruitment email. So it may be that the response rate for career scientists and engineers was somewhat higher than 36.5 percent. Although the response rate does not appear to be out of range with similar studies of this type, there are likely to be some forms of non-response bias. In short, it is acknowledged that there are generalizability and external validity limitations of this research.

8.2 Limitations of Modeling and Model Assumptions

8.2.1 Model Assumptions

Basic model assumptions—such as relative conformance with normality, zero mean value and constant variance of the disturbance term, the lack of correlation between the explanatory variables and the disturbance, and correct model specification—are problematic when attempting to model a phenomenon (i.e., productivity) that is highly skewed by nature. There are strong outlying cases in the data that are not due to data quality problems but are due to the existence of a minority of the scientific and engineering population that is very different from the majority.

Tobit models, while useful in dealing with censored data, are sensitive to model misspecification. There is surely model misspecification in this study (as there is in many studies) due to the limitation of available data and the difficulty in measuring, operationalizing, and controlling for variables such as innate motivation, human capital, and social capital.

In addition, Tobit models assume specific forms of censoring such as censoring from below at zero or above at some other threshold value. In this study, there is likely to be both forms of censoring. While the major censoring is censoring from below at the threshold level of zero, the Tobit analysis cannot correct for multiple forms of censoring when the threshold values are not know or are variant across observations (i.e., cases).

Poisson models are appropriate for modeling count data where there are a non-trivial number of zeros among the counts and where large counts are relatively rare. However, Poisson models assume that the mean of the distribution is equal to its variance. In cases, where the variance exceeds the mean, so called over dispersion, standard errors are biased toward zero and significance tests and p-values cannot be relied upon. There may be some overdispersion in this dataset. One solution is to model count data as a negative binomial distribution via maximum likelihood estimation, which will remain as an opportunity for future research.

One possible solution to the above parametric modeling assumptions is to employ non-parametric methods to avoid the above problems. Neural Network models are non-parametric in form and are believed to benign to parametric assumptions and "noisy" (e.g., missing values or measurement error) data. Because input neurons are weighted and adjust through iteration in an attempt to approximate their targets with minimum error, the weights on the hidden neuron layers can be used to approximate the effects of various inputs on the target. Neural Network analyses can be trained and validated on subsamples of the data, which has been done in this study, to avoid the problem of over-training (i.e., over valuing the observed data at the cost of minimizing generalization). Despite their advantages, Neural Network models do have problems in generalizability unless they are tested against samples of new data. In addition, because they are highly dynamic, model interpretation and coefficient interpretation are not as straight forward as is the case in parametric models.

8.2.2 Appropriate Modeling Techniques

None of the available non-Bayesian, parametric models, which are designed to test the effect of explanatory variables on a dependent variable, truly capture the essence of measurement and analysis of dynamic career and productivity patterns. In this study several variables (e.g., the homogeny index, first job sector, proportion of career years in industry jobs) that summarize these patterns into a single number have been included. Yet, it seems to me that this does not do justice to the problem as posed. The career homogeny index is based on the logic of Hidden Markov models—which use strings of conditional probabilities of observed states and an unobserved "hidden" chain to classify patterns (discussed in Section 8.5)—may hold promise for future research in this area.

8.3 Validity of Measures

8.3.1 Diversity of Job Pattern

The diversity of the job pattern in terms of the homogeny index and intersectoral career changes is limited (in its construct validity) as a measure or proxy for the social and human capital building experiences these environmental changes may embody. So it may not be the number or type of these career transformations that matter but the actual quality and character of the experience provided through these job transformations. The human and social capital building opportunities that these job changes may or may not represent is likely to be a better and more sensitive indicator, but one that is fraught with measurement difficulties and cannot easily be extracted from the CV. As a result, these variables are proxy variables in this respect.

8.3.2 Publication and Patent Rates as a Measure of Productivity

Likewise, publication and patent counts and rates are relatively crude measures of productivity. Citations to publications may be a better overall measure of the contribution that a scientist or engineer has made to the knowledge base. While publication counts are a measure of "work" productivity (units of output over time) as opposed to the perhaps more relevant concept of "knowledge productivity," which would be better measured by citation counts.

While citation analysis scholars have thoroughly inventoried the shortcomings of citations as measures of quality and peer recognition of scientific work (see, for example, Narin and Hamilton, 1996; Narin, 1994; Narin, Olivastro, and Stevens, 1994; Cozzens, 1989; MacRoberts and MacRoberts, 1986; Narin, 1976), citation counts generally serve as a good measure of "usefulness" of the scientific and engineering research. The Cole brothers found, in a 1971 paper that reviewed empirical studies of citations as a measure of quality, that they generally correlate quite highly with peer assessments of quality and just about every other measure they tested (Cole and Cole, 1971).

The use of citation analysis was explored for this study but due to the enormous financial cost of purchasing these data and the enormously labor intensive process of matching such data to career records, such analysis was impossible. Even use of the so called Journal Impact Factor, the mean number of citations to articles in a given journal in a given year, would have required the recoding of the 65,320 publications of the scientists and engineers in the RVM dataset. Moreover, the variance in the impact factor over time tends to be high in some journals versus others. Such a proof-of-concept study where the impact factor is used as a weight on a subsample of the cases may be possible in future research.

8.4 Limitation of Findings

The above limitations in the research design and methodology, affect the limitations of the findings. However, there are several other limitations. First, the nature of the association among explanatory and dependent variables (i.e., internal validity) may be questioned like all statistical models based on correlation. There is likely to be some form of reciprocal correlation between productivity and other variables. For example, grant awards may boost productivity, while productivity may increase the likelihood of receiving grant awards. Likewise, job changes may affect productivity or vice versa, although the extant research findings discussed in Chapter 2 and those discussed in Chapter 5 (Section 5.4) suggest that productivity is affected positively by the job move.

Finally, it may be the case that holding multiple jobs concurrently (such as Department Chair, Endowed Chair, or Director of a research center) positively affects productivity. On the other hand, these concurrently held jobs provide affordances not available to other researchers, such as a greater number of graduate students, postdoctoral research associates, and research scientists, who may do much of the paper authoring. However, much of the research related to this topic (as discussed in Chapter 2) suggests that productivity only weakly affects prestige of next job and that productivity increases after the attainment of the next job post. Likewise in examining job transformations to and from industry, as part of this study, productivity increased after the job move (regardless of job sector). However, this may be an artifact of the fact that job moves between academia and industry may be more likely to occur at the career stage when productivity (in general) is on the increase. Nonetheless, there may be some cause for concern about non-recursive relationships such as endogeneity within the explanatory variables.

8.5 Opportunities for Future Research

There are several possible avenues for productive future research using the RVM dataset and through the collection of supplementary data from the CVs and other sources.

8.5.1 Social Network Analyses

One of the limitations of this research is the difficulty in measuring human and social capital. While human capital may be especially difficult to measure, it is possible to do a better job in measuring social capital, at least as it is embodied through social networks. Social network analyses (Wasserman and Faust, 1994; Scott, 2000) could be constructed for some subsample of the RVM dataset, perhaps for scientists and engineers affiliated with a few related centers, through coauthorship patterns. With social network analyses, the centrality of key researchers can be identified as can the relative density, sparseness, and diversity of publication collaboration for all researchers. Measures of these concepts can be constructed to examine the effects of sustained or diverse collaborative publishing on productivity. Although it would be relatively simple to do this for several of the research centers and their affiliated researchers in the RVM dataset, social networks can become intractable and labor intensive as the sample size increases.

8.5.2 Hidden Markov Models

Hidden Markov Models (HMMs) are used in pattern recognition problems of various sorts but were first (and remain most commonly) used in artificial speech recognition models and related software development. In this context, HMMs could be used to identify and classify job patterns, which could then be examined for productivity differences.

The HMM is a multi-state conditional probabilistic model. Each state-transition generates a set of probabilities (i.e., the transition matrix). In conditional HMMs[43] the probability that any specific state (e.g., job position) will follow the current state can be estimated using forward and backward algorithms. HMM usually assumes that the probability of the next state is dependent on the prior state at time t-1. This is called a first order HMM. A second order HMM depends upon the two previous states and so forth (Bengio, 1999).

[43] In homogenous HMMs the transition matrix is assumed to be static for any given state-transition at any stage of the chain. In conditional HMMs, the transition matrix is allowed to vary depending on location in the chain.

One problem with Markov Chains of higher order is that they tend to become computationally intensive as the order becomes arbitrarily large. The hidden component in HMM is thus added as a continually adjusting term that summarizes the effect of previous states.

The joint probability distribution of any given chain is determined by three sets of probabilities—the initial state probability (which is given for each state value ($P(q_1)$)), the transitional probability (which is $P(q_t \mid q_{t-1})$, and the emission probability ($P(y_t \mid q_t)$). What is estimated is the probability of observing a particular sequence given any specific emission value (y_t) (Bengio, 1999).

HMMs may prove effective in modeling productivity as a function of career states and sequences, where the states are defined as a researcher's job positions over time. One important and attractive benefit of using HMMs to model job patterns and productivity is that they are robust to chains of differing length. This is not the case with most parametric models used in the social sciences where a job variable (say the 10th) job is treated as missing data for all cases where fewer than 10 jobs are observed. The possible analyses could suggest various patterns of job sequences that are more or less associated with higher levels of productivity than others. Unfortunately, currently, most HMM software are written with a specific (usually speech recognition) application in mind. Yet the method holds promise in embodying the entire career pattern—not just a job or a summary measure of job types—as the unit of analysis as discussed in Chapter 1.

8.5.3 Further Refinement and Validation of Neural Network Models

Another possible future research project would involve further investigation of the Neural Network approach. Although I presented several preliminary models in this dissertation, further refinement and validation using new data may be warranted. It would be possible to make use of new data (currently being collected in RVM Phase III) to test the cross sample validity of the models and to assess the differences in findings generated under the new data.

Moreover, there are other types of Neural Network models, which may hold promise for future research. One such model, the Kohonen Self-Organizing Maps, is particularly useful in pattern recognition because inputs are searched for patterns, which are then associated into clusters. There are also several models, such as Adaptive Time-Delay Neural Networks, that might hold promise in modeling time-dependent, chronological data such as is the case in career patterns (Garson, 1998).

8.5.4 The Use of Qualitative Methods

In addition to its value as a stand-alone source of data, a great advantage of CV data is that it can be used in conjunction with other sources of data. Now that much of the CV data has been analyzed and the findings in this study have been presented, it may be possible to conduct several qualitative studies that further refine and expand on the findings of this study. One such research project would include indepth interviewing of the publication and productivity stars compared with others in order to improve understanding of what makes them different from each other and from the non-stars. In addition, the finding of no relationship between productivity and having had a postdoctoral position deservers further exploration in a qualitative, indepth way.

CHAPTER 9

CONCLUSIONS

In this concluding chapter, I will first summarize the main findings in section 9.1. Then, in section 9.2, I will revisit the findings briefly and discuss the knowledge implications of this study in reference to the literature on this topic. Next, I will discuss the implications this work has for policy and to research evaluation methodology (section 9.3) as promised in Chapter 1.

In short, the key findings are as follows:

1. There is substantial intersectoral career pattern diversity, research funding, and commercially oriented output (such as patents) among researchers in the RVM dataset. The extant literature base on academic careers has not adequately recognized this. As research centers continue to become an important component of the “academic” context, understanding the contemporary academic career becomes more important to research evaluation and science and technology policymaking (see discussion in section 9.1.1).

2. The extant literature that relates prestige factors to productivity enhancement may be misattributing cause and effect. This research demonstrates that publication rates increase after job changes from industry to academia as well as from academia to industry. It is possible that this is due to human and social capital factors or due to a correlation between periods when job changing is likely to occur with periods where productivity is naturally on the rise (see discussion in section 9.1.2).

3. There is evidence that having had a postdoctoral research position is negatively related to overall career productivity rates. This may be due to several factors (see discussion in section 9.1.3), direct or spurious. In either case, there is a substantial public investment in these positions and further investigation is thus warranted.

4. There is reason to believe that publication and patent stars exhibit different career patterns, the latter focused more on relationships with industry. But, there is also reason to believe that publication productivity is related to higher degrees of career homogeny (making more typical career choices) than patent productivity. This suggests an important policy question worthy of further investigation. The NSF approach to guiding centers toward commercial outputs is to encourage them to seek industry representation on their advisory boards. If NSF desires a greater commercial orientation of these centers, further qualitative and quantitative research and evaluation may consider these human capital and career factors in understanding the role these centers play in the triple helix. (see section 9.1.4 and section 9.2).

9.1 Summary of Main Findings

9.1.1 There is More Career Diversity in Academia Than the Literature May Imply

The literature base on the academic workforce has looked narrowly at academic job changes. The result may leave a misleading picture of the career patterns of productive scientists and engineers. As research centers become a substantial and recognized component of federal science policy structures, this research reveals a picture of an important component of the academic research community that is different from that portrayed in the literature.

First, researchers in the RVM data set had substantial experience working in industry and government. On average, nearly one in six of their total jobs positions were industry jobs and one in eight of their career years were spent in industry jobs. Approximately half had one or more jobs in industry. One-quarter had worked for government. And, nearly half began their careers in non-academic jobs—for 33 percent the first job was in industry and for 15 percent the first job was government. An average of 24 percent of the grants awarded to researchers in the RVM dataset came from industry.

Second, it is commonly thought that it is hard to enter academia later in a researcher's career. Nearly one in ten took their first academic job 10 or more years into their career, and more than one in five of the RVM researchers took their first academic job 5 or more years into their career.

These latter statistics suggest that there is perhaps substantially more intersectoral linkages than may be commonly perceived (at least for this dataset of researchers working in academic research centers). And, generally speaking, these findings hold even among publication stars.

Third, more than 40 percent of RVM researchers had one or more patents and it is clear that the career patterns associated with patent productivity are different than those associated with publication productivity—namely more industry jobs and funding and less homogenous career patterns. So clearly patent rate is an important measure of research productivity but has been largely ignored in the extant literature.

These three findings when taken together suggest that the bulk of the literature on academic productivity is missing important communities within the academic workforce and may be improperly, and perhaps inadvertently, painting a picture of the academic career that is more or less monolithically academic in nature, sequence, and pattern.

9.1.2 Productivity and Prestige

To test the assertion that intersectoral changes in jobs affects overall productivity, all of the transformations from academic jobs to industrial jobs and vice versa were identified and the mean number of publications was calculated for the five years before and after the transformation. These means were then summed and averaged over all job transformations made by all scientists and engineers between academia and industry. As seen in Table 16, the mean number of publications for the five-year period preceding a job transformation *from industry to academia* (for all scientists and engineers who made this transition) was 1.5 publications per year compared with a mean of 2.6 publications after the move—a 1.1 increase in the mean number (and a statistically significant difference at the .05 level). Thus, the average productivity of scientists and engineers increased after a job transformation from industry to academia.

A similar analysis was performed for all job transformations made by the scientists and engineers *from academia to industry*. For the five-year period preceding the move, the mean number of publications overall was 1.8; for the five-year time interval following the transformational move to industry, the mean was 2.6—an increase of .8 publications per year (also a statistically significant difference). Thus, the mean number of publications for the five-year period of time following a move from academia to industry also increased.

However, the prestige literature (and parts of the accumulative advantages literature) suggests that productivity increases *after* a job change to a more prestigious institution or department. This served as a basis for the accumulative advantages hypothesis. However, my analysis demonstrates that job changes between academia and industry also result in higher post-job-move productivity—even when the move is to industry where publication productivity is less highly rewarded.

This suggests that prestige may have nothing to do with the productivity boost. Although making a job change may have important effects on productivity, perhaps due to new social and human capital endowments, an alternative interpretation is likely. Making a job change may be correlated with the career stage when productivity rates are naturally increasing. For example, it may be the case that researchers in the early stages of their careers are more likely to make job changes than those in later stages and that this is precisely the same time period when researchers' productivity is on the rise. Finally, the fact that job moves to industry also resulted in higher publication productivity, suggests that better measures of human and social capital as applied to intersectoral job changes may indeed reveal further insight into this finding.

9.1.3 Do Postdoctoral Researchers Have Lower Career Productivity Rates?

Across all of the models, the effect of having had a postdoctoral research position appears to be negative (see Figure 16).

Statistical Modeling Technique	Finding
Tobit	The coefficient on the postdoctoral variable was negative but not statistically significant. -0.31 for publication rate; -0.06 for patent rate
Poisson	The coefficient on the postdoctoral variable was negative and large but not statistically significant (-0.16)
Stars	32 percent of non-stars compared with 22 percent of patent stars held postdoctoral positions, compared with 24 percent for publication and combination stars
Neural Networks	The model estimated researchers without postdoctoral positions had a publication rate more than twice that of those who had held a postdoctoral job. Those with no postdoctoral job had a patent rate more than 5 times those who had postdoctoral jobs.
Descriptive Statistics—Publications	Those who had no postdoctoral job averaged 3.6 publications per year, those with one postdoctoral job averaged 3.2 publications per year, and those with more than one postdoctoral job averaged 2.7 publications per year.
Descriptive Statistics—Patents	Those who had no postdoctoral job averaged 0.16 patents per year, those with one postdoctoral job averaged 0.10 patents per year, and those with more than one postdoctoral job averaged 0.02 patents per year.

Figure 16. Summary of Findings on Postdoctoral Researchers Across All Models and Analyses

This finding, if corroborated by studies on this topic, suggests that postdoctoral positions may be ineffective in terms of helping researchers to establish productive research careers over the long term. First, it is possible that they have shorter term impacts on productivity or that they have other positive effects such as providing financial access to scientific and technical education and careers. Moreover, postdoctoral positions are often located in the academic sector where patent productivity is not rewarded as highly as publication productivity.

Second, it may be the case that postdoctoral associates are vi ewed as research “labor” and are not credited with break through discoveries because their work is dominated by so called intellectual “mopping up” duties.

Third, it may be the case that postdoctoral research positions have become more common over time so this finding may be spuriously correlated with age cohort. However, this explanation does not seem to be borne out in an analysis of the frequency of postdoctoral research positions by age cohort. There also seems to be no substantial difference in this finding by disciplinary field.

Finally, postdoctoral positions may have a form of career “holding pattern” effect where top graduates are sought after for assistant professorships and others are left waiting in the career queue. Yet all of these possibilities are policy relevant due to the large public expenditures on supporting such positions. As a result, this finding deserves further exploration and may require a more thorough study involving the use of theories from the science of learning and better empirical measures.

9.1.4 What Is the Relationship Between Job Sequences and Productivity?

One approach I took to examining the sequences of jobs and their relationship to productivity was to examine each job change as a conditional probability (e.g., what is the likelihood given a research is in job A that he or she proceeds to job B). All 5,490 job transformations in the RVM dataset were used to build these relative frequencies. For each researcher, the chain of job transformation probabilities was summed and divided by the total number of jobs for that researcher. The result was a variable (career homogeny) that gives a relative perspective on whether a researcher has had a relatively typical or atypical career sequence.

Among the productivity stars, it appears that combination stars (the most highly productive researchers) had the highest career homogeny (i.e., more typical career paths), followed by publication stars, followed by non-stars. Patent stars had the lowest career homogeny indexes (see Figure 17).

	Means			
Variable	Publication star	Patent star	Combination star	Non-star
Career homogeny index	*15.3	§12.3	+16.8	14.7

Notes: * Indicates a statistically significant difference in means between publication star and patent star. § Indicates a statistically significant difference in means with non-star. + Indicates a statistically significant difference between patent star and combination star.

Figure 17. Summary of Career Homogeny Indexes Among Productivity Stars

This suggests two observations. First, it appears that typical career patterns are more strongly associated with high publication rates than is true for patent rates. Second, at least as estimated in the Neural Network models (see Figures 18 and 19), it appears that the strongest productivity benefits of career homogeny are not realized except for a relatively elite group whose career homogeny rates are well above those achieved by the mean values of the productivity stars.

Figure 18. Estimated Relationship Between Career Homogeny and Publication Rate

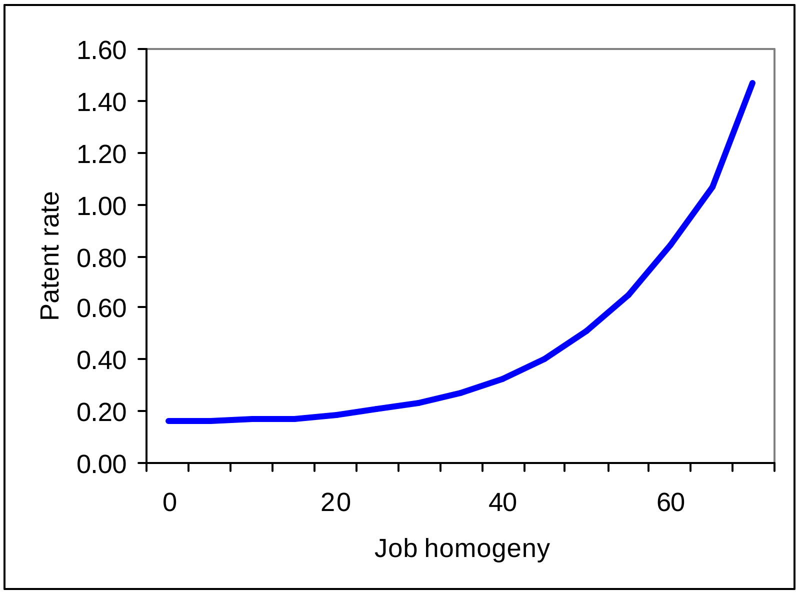

Figure 19. Estimated Relationship Between Career Homogeny and Publication Rate

Finally, an examination across researchers with relatively homogenous and inhomogenous career patterns reveals that those with the highest career homogeny indexes often contain an unbroken or relatively unbroken direct sequence of assistant professor, associate professor, full professor. Often it does not seem important what job or jobs precede or follow this particular sequence, it is more important that the sequence is there and relatively self contained from job patterns before or after.

9.2 Other Findings Relevant to Policy and Theory

First, one of the main conclusions of this study in terms of its contribution to the literature on academic career paths is its focus on the use of patents as a knowledge productivity indicator. Through the focus on both publication and patent productivity, I believe that this research has contributed to our understanding of productivity and how it may manifest itself differently in terms of career paths. This conclusion is justified through the findings that demonstrate different variables are associated with publication and patent productivity, respectively. So by ignoring patent productivity previous studies may have ignored important aspects of academic career paths and important measures of knowledge productivity.

If NSF wishes to increase the commercial orientation of its centers and their relevance to industry, it should consider further qualitative and quantitative research and evaluation to examine these human capital and career factors in understanding the role these centers play in the triple helix (of academia, industry, and government).

Second, although there are conflicting results across models, generally speaking, publication productivity seems to be characterized by less employment in industry and relatively greater involvement in government jobs, greater career homogeny (at least for the publication stars), more job appointments per career year, and a higher level of grants per year. Patent productivity, on the other hand, seems to be characterized by greater involvement with industry in terms of jobs and grant support and less involvement with government jobs, and lower levels of career homogeny (at least among the patent stars), and perhaps fewer job appointments per career year (although this latter difference is not statistically significant).

Third, this research provides an opportunity to reflect on the prestige and accumulative advantages hypothesis well rehearsed in the literature. It may be that the emphasis is misplaced. It may well be that a progression of career "advantages" are at work, along with the prestige and recognition that goes with them, but it is also possibly the case that these are human and social capital advantages not just visibility and prestige advantages. These individuals seemed to have worked and progressed in their careers to become center directors, members of advisory committees of importance, visiting professors, editorial boards members—all of these can be considered human and social capital building opportunities.

Fourth, in terms of the actual hypotheses, there is little evidence to support the education and human resources hypothesis. Precocity is only weakly related to productivity and seems only to matter in the extremes as evidenced in the neural models. So highly precocious graduate students may indicate high future productivity but precocity at low to moderate values seems not to make a difference. As mentioned above, postdoctoral positions were negatively related to overall career productivity.

As for the diversity hypothesis and the homogeny hypothesis, the evidence suggests that the homogeny hypothesis is better supported. However, there is quite a bit of intersectoral diversity in the academic careers of scientists and engineers in the dataset, perhaps more than I would have originally predicted. The evidence in support of the diversity hypothesis is generally weak with the exception of the finding that lesser degrees of career homogeny is associated with patent stars as compared to the other groups. As a result, it may be better to ask at what point or in what way do industrial and governmental experiences help or hinder academic productivity. Perhaps, again, this is the subject of qualitative work and better measures of social and human capital.

9.3 Policy Implications of This Work

First, the formation of S&T human capital theory (Bozeman, Dietz, and Gaughan, 2001) has important policy implications for how knowledge is created and how research may be evaluated. This theory led to the conceptualization of the career as an important unit of analysis and the CV as an appropriate and useful source of data. This, in turn, led to early work on the methods of coding and analyzing CV data (Dietz, et al., 2000). The use of CVs, although not without shortcomings, holds a wealth of possibilities in assessing research evaluation and science policies. As government accountability demands continue to grow (Bozeman and Melkers, 1993; Government Performance and Results Act of 1993), the technology of research evaluation must grow accordingly. The detailed treatment of methodology is an advance in the knowledge base on this topic. This advance is not only methodological, it has led to important insights into the diversity of the scientific and engineering careers discussed above, which are not well addressed in the academic career literature.

In addition, I have shown that the statistical models presented in this dissertation—including Tobit and Poisson models and Neural Network models—each have advantages based on modeling assumptions for the analysis of CV data and productivity data. So the goal of proof-of-concept evidence for the method and the statistical methodologies to support it has been met. Thus, theory has led to new forms of data collection, which have led to new forms of data analysis, and a reconceptualization of the academic career.

Second, there are several policy implications deserving of further research. More research into the effects of postdoctoral appointments is warranted. From a human resource point of view, one would expect these experiences to be crucial in shaping a scientist's or engineer's research identity, frame of reference and theoretical perspectives, "nose" for interesting problems, and cognitive and craft skills. However, they did not appear to be associated with higher productivity. If further investigation should corroborate this finding, there are a number of interesting policy questions about why this is the case, and it may appropriate to explore more productive alternatives for early career development.

Third, if government policy chooses to focus on the commercial productivity of academic institutions as seems to be the policy trend then there is reason to believe that the solution may be human capital in nature. The hiring of researchers with industrial job experience and, perhaps, visiting positions and exchanges with industry may be a productive means of boosting the commercially-relevant innovation of universities. NSF's GOALI program may be one prototype worthy of evaluation using methods such as those used in this dissertation. In fact, an interesting study would be to examine this in the context of countries in the developing world that seek to boost the economic payoff of investments in academic institutions[44].

Fourth, of the scores of centers involved in this research, only a few seem to stand out as places of higher productivity and it is not clear what role the center plays in that, although this may be due to relatively lower response rates in some centers. An interesting study would be to focus on these centers and do some comparative analyses to get a richer and better understanding of the internal validity of the variables used in this research.

Finally, among the most important policy-relevant future research directions would be to develop a better measure of social capital and also to work from psychological theory to formulate better measures of innate motivation and human capital. In the context of richer work within the

[44] It is acknowledged that the relationship between patenting activity and economic growth is complex. But it is an indicator of potential economic value of a discovery.

centers, this would enable the researcher to make claims about the relative importance of social, organizational, or environmental, and personality and psychological factors on productivity.

APPENDIX A: Tables

Table 1. Center Titles and Response Rates

Center code #	Center	CVs requested	CVs obtained	Bad address	Response Rate
1	Membrane Applied Science and Technology (University of Colorado)	8	3		0.375
2	Neuromorphic Systems Engineering, California Institute of Technology (ERC)	206	53	13	0.275
4	Biotechnology Process Engineering Center, Massachusetts Institute of Technology (ERC)	8	6		0.750
5	Engineered Biomaterials, University of Washington (ERC)	36	14	7	0.483
6	Environmentally Benign Semiconductor Manufacturing, University of Arizona (ERC)	97	30	3	0.319
7	Innovation In Product Development, Massachusetts Institute of Technology (ERC)	141	27	18	0.220
8	Reconfigurable Machining Systems, University of Michigan (ERC)	31	14	3	0.500
9	Computational Field Simulation, Mississippi State University (ERC)	57	12	2	0.218
10	Data Storage Systems Center, Carnegie Mellon University (ERC)	28	17	1	0.630
11	Telecommunications Research, Columbia University (ERC)	44	10	2	0.238
12	Low Cost Electronic Packaging, Georgia Institute of Technology (ERC)	47	17	3	0.386
13	Compound Semiconductor Microelectronics, University of Illinois (ERC)	92	5	29	0.079
14	Integrated Media Systems Center, University of Southern California (ERC)	25	17	1	0.708
15	Advanced Technology for Large Structural Systems, Lehigh University (ERC)	7	2		0.286
16	Particle Science and Technology, University of Florida (ERC)	39	23	4	0.657
18	Advanced Engineering Fibers and Films, Clemson University (ERC)	25	9	1	0.375
19	Advanced Combustion Engineering Research Center, Brigham Young University (ERC)	15	10		0.667
20	Pacific Earthquake Engineering Research (PEER) Center (ERC)	119	42	6	0.372
21	Mid-America Earthquake (MAE) Center (ERC)	31	13		0.419
22	Glass Research (Alfred University)	21	10	3	0.556

Table 1 (cont'd).

24	Advanced Steel Processing and Products Research (Colorado School of Mines)	8	4		0.500
25	Coatings Research (Eastern Michigan University)	8	4	1	0.571
27	Micro-engineered Ceramics (University of New Mexico)	12	6	2	0.600
29	Engineering Tribology (Northwestern University)	15	9		0.600
31	Advanced Polymer and Composite Engineering (Ohio State University)	14	8		0.571
32	Particulate Materials (Pennsylvania State University)	12	6		0.500
33	Dielectrics (Pennsylvania State University)	36	13	2	0.382
34	Biological Surface Science (SUNY at Buffalo/Alfred University/University of Memphis/University of Miami	5	3		0.600
36	Advanced Control of Energy and Power Systems (Arizona State University)	53	19	3	0.380
37	The Built Environment (University of California, Berkeley)	21	14		0.667
38	Material Handling Logistics Institute (Georgia Institute of Technology)	41	23	5	0.639
40	Nondestructive Evaluation (Iowa State University)	25	8		0.320
42	Quality and Reliability Engineering (Rutgers University)	10	6		0.600
43	Measurement and Control Engineering (University of Tennessee)	12	7		0.583
45	Integrated Pest Management (North Carolina State University)	4	1		0.250
46	Management Information (University of Arizona)	33	5	2	0.161
49	Research of Information Technology and Organizations (University of California, Irvine)	43	15	2	0.366
50	Ultra-High Speed Integrated Circuits and Systems (University of California, San Diego)	21	9		0.429
51	Sensors and Actuators (University of California, Berkeley)	136	27	13	0.220
53	Wireless Information Networks (Rutgers University)	43	14		0.326
54	Advanced Electronic Materials, Devices and Systems (University of Texas at Arlington)	55	9	6	0.184
55	Design of Analog/Digital Integrated Circuits (Washington State University/University of Washington/Oregon State University/SUNY Stony Brook)	15	7	1	0.500

Table 1 (cont'd).

56	Software Engineering (University of Florida/Purdue University/University of Oregon/West Virginia University)	29	13	1	0.464
57	Advanced Computing and Communication (North Carolina State University /Duke University)	22	8	1	0.381
58	Advanced Air Conditioning and Refrigeration (University of Illinois, Urbana)	72	24	12	0.400
59	Process Analytical Chemistry (University of Washington)	51	20	2	0.408
60	Behavioral Neuroscience (Emory)	79	46	5	0.622
61	Microbial Ecology (Michigan State)	45	19	3	0.452
62	Graphics and Visualization Center – (Brown)	31	15	1	0.500
63	Nanobiotechnology Center (Cornell)	24	10	2	0.455
64	Analysis and Prediction of Storms (Oklahoma)	32	10	9	0.435
66	Advanced Liquid Crystalline Optical Materials (Case Western Reserve U.)	29	18	1	0.643
67	Environmentally Responsible Solvents and Processes	119	50	3	0.431
68	High-Performance Polymeric Adhesives and Composites (Virginia Polytechnic Institute and State University)	19	6	2	0.353
69	Particle Astrophysics	131	24	40	0.264
70	Photoinduced Charge Transfer	75	11	32	0.256
71	Quantized Electronic Structures	65	12	20	0.267
73	Electronic Imaging Systems (CEIS)	27	11	3	0.458
76	Optoelectronic Computing Systems Center (ERC)(Uni. Of Colorado-Boulder)	6	4		0.667
76	Optoelectronic Computing Systems Center (ERC)(Uni. Of Colorado-Boulder)	69	28		0.406
78	Plant Sensory Systems Network (ERC)(Ohio State Univ.)	13	12		0.923
81	Institute for Research in Cognitive Science (STC)(U.Penn)	7	7		1.000
82	Biofilm Engineering (ERC)(Montana State University)	54	14		0.259
83	Computation and Neural Systems (ERC) (Caltech)	91	13		0.143
84	Light Microscope Imaging and Biotechnology (STC)(Carnegie Mellon)	66	13	3	0.206
85	Ultrafast Optical Science (U. Michigan) (STC)	94	45		0.479
86	Clouds Chemistry and Climate (UCSD) (STC)	2	2		1.000
87	Biological Timing (U. VA) (STC)	74	32		0.432
89	Synthesis, Growth, and Analysis of Electronic Materials (U.Texas) (STC)	72	28		0.389
90	James R. Macdonald Laboratory (Kansas State University) (DOE)	22	2		0.091
92	Interconnect Focus Center	64	48		0.750

Table 1 (cont'd).

3/ctrref	Emerging Cardiovascular Technologies, Duke University (ERC)	2	2		1.000
17/ctrref	Advanced Electronic Materials Processing, North Carolina State University	50	2	16	0.059
23/ctrref	Steel Making Research (Carnegie Mellon University)	5	2	1	0.500
26/ctrref	Polymer Biodegradation (University of Massachusetts)	4	1	2	0.500
28/ctrref	Energetic Materials (New Mexico Institute of Mining and Technology)	7	2	1	0.333
30/ctrref	Corrosion in Multiphase Systems (Ohio University)	15	3		0.200
35/ctrref	Ergonomics (Texas A & M University)	33	5	17	0.313
39/ctrref	Machine-Tool Systems (University of Illinois)	5	1	1	0.250
41/ctrref	Silicon Wafer Engineering and Defect Science (North Carolina State University)	5	1		0.200
44/ctrref	Hazardous and Toxic Management (New Jersey Institute of Technology/Tufts University)	8	1	3	0.200
52/ctrref	Study of Wireless Electromagnetic (University of Oklahoma)	3	2		0.667
65/ctrref	High Pressure Research (SUNY- Stony Brook)	13	2	4	0.222
72/ctrref	Intelligent Information Retrieval(CIIR)	7	2		0.286
74/ctrref	Low Power Electronics	5	2		0.400
75/ctrref	Advanced Friction Studies	6	2		0.333
79/ctrref	Engineering Design Research Center (ERC) (Carnegie Mellon)	5	4		0.800
80/ctrref	Institute for Systems Research (Uni. Of Maryland)	15	2		0.133
91/ctrref	Microelectronics Research Center - Spring 2000	183	64		0.350
*	Engineering of Living Tissues, Georgia Tech/Emory University (ERC)	5	1		0.200
*	Collaborative Manufacturing, Purdue University (ERC)	2	0		0.000
*	Pharmaceutical Processing Research (Purdue University)	2	0		0.000
*	Web Handling (Oklahoma State University)	1	0		0.000
*	Aseptic Processing and Packaging Studies (North Carolina State University)	1	0		0.000
*	Optoelectronic Devices, Interconnects, and Packaging (University Maryland)	7	1	1	0.167
*	Sustainability of Semi-Arid Hydrology and Riparian Areas	1	0	1	
Total		**3573**	**1198**	**320**	**0.37**

* indicates no response or CV was eliminated because of status as student or administrative support staff

Response Rate = Number of obtained CVs / (Number of CVs requested - Unreachable)

Table 2. Intercoder Reliability and Time of Coding for 10 CVs (Pilot Study)

Curriculum Vita	Rs(i) for Coding Experiment 1	Coding Time (in minutes) Experiment 1	Rs(i) for Coding Experiment 2[†]	Coding Time (in minutes) Experiment 2
1	.897	23.0	.938	18.0
2	.797	31.0	.765	21.0
3	.651	27.8	.881	18.4
4	.839	23.4	.709	14.4
5	.868	24.0	.792	15.6
6	.608	15.0	.830	16.0
7	.756	30.4	.830	13.8
8	.800	19.4	.630	21.8
9	.728	19.8	.832	18.0
10	.714	22.2	.849	10.0
Mean	.766	23.6	.805	16.7
Std. Dev	.090	5.02	.088	3.51

Note: Rs(i) stands for resume intercoder reliability.

[†] Coding experiment 2 used the same coders with different CVs and an improved coding protocol.

Table 3. Intercoder Reliability for Coding 37 Items (Pilot Study)

Item #	Rs(i) Coding Trial 1	Rs(i) Coding Trial 2†	Item Name
1	.933	.682	CV version is full or partial
2	.933	1.000	Sex of respondent
3	.960	1.000	Year of birth
4	.920	.800	National origin
5	.920	.780	Citizenship
6	.880	1.000	Degree type of first degree
7	.690	.840	Degree field of first degree
8	.860	.880	Degree type of second degree
9	.880	.670	Degree field of second degree
10	.960	.960	Degree type of third degree
11	.810	.800	Degree field of third degree
12	1.000	1.000	Degree type of fourth degree
13	1.000	1.000	Degree field of fourth degree
14	1.000	1.000	Degree type of fifth degree
15	1.000	1.000	Degree field of fifth degree
16	* .560	* .570	Job title of first job
17	* .490	* .490	Job title of second job
18	.610	.770	Job title of third job
19	.680	.680	Job title of fourth job
20	.600	.410	Job title of fifth job
21	.740	1.000	Publication type of most recent pub.
22	.790	.960	Publication type of second most recent pub.
23	.630	.860	Publication type of third most recent pub.
24	* .550	.860	Publication type of fourth most recent pub.
25	.880	.760	Publication type of fifth most recent pub.
26	.620	.780	Dollar amount of first grant or contract
27	* .570	* .520	Funding source of first grant or contract
28	.630	.760	Dollar amount of second grant or contract
29	* .580	.680	Funding source of second grant or contract
30	.690	.760	Dollar amount of third grant or contract
31	* .560	.710	Funding source of third grant or contract
32	.690	.760	Dollar amount of fourth grant or contract
33	.660	.620	Funding source of fourth grant or contract
34	.660	.780	Dollar amount of fifth grant or contract
35	* .580	.690	Funding source of fifth grant or contract
36	.893	.960	Year of first patent
37	.933	1.000	Was first patent licensed or sold?
Mean	.766	.805	
Std. Dev.	.164	.163	

Notes: Rs(i) stands for item intercoder reliability.
† Coding experiment 2 used the same coders with different CVs and an improved coding protocol.

Table 4. Descriptive Statistics

Variable group and name	N	Minimum	Maximum	Mean	Std. Dev.
Demographics					
Sex	950	0.00	1.00	0.878	0.328
Year of birth	328	1923	1974	1950	10.728
Education					
Year of bachelor's degree	891	1940	1997	1976	10.984
Year of master's degree	668	1942	1999	1979	10.979
Year of doctoral degree	913	1943	2000	1982	11.467
Year of other degree	130	1952	1999	1982	11.445
Doctorate field in biological sciences? (1=Yes, 0=No)	956	0.00	1.00	0.105	0.306
Doctoral field in computer Science	956	0.00	1.00	0.046	0.210
Doctoral field in engineering	956	0.00	1.00	0.452	0.498
Doctorate field in physical sciences	956	0.00	1.00	0.283	0.451
Doctorate in other field	956	0.00	1.00	0.114	0.318
Career age and age cohort					
Career length (years since doctorate)	951	1.00	58.00	18.145	11.492
Doctorate granted before 1972? (1=Yes, 0=No)	951	0.00	1.00	0.196	0.397
Doctorate granted 1972-1980	951	0.00	1.00	0.187	0.390
Doctorate granted 1981-1987	951	0.00	1.00	0.162	0.369
Doctorate granted 1988-1993	951	0.00	1.00	0.218	0.413
Doctorate granted 1994-2000	951	0.00	1.00	0.201	0.401
Publications					
Total number of publications	933	0.00	628.00	75.290	92.190
Total publications/career Length	929	0.00	34.00	3.986	3.530
Precocity (see notes)	956	0.00	17.00	3.160	3.454
Patents					
Total number of patents	956	0.00	141.00	2.704	8.272
Total patents/career length	953	0.00	5.22	0.143	0.387
Jobs					
Number of job positions, total [total jobs]	956	0.00	26.00	6.712	3.740
Number of job institutions	956	1.00	11.00	3.288	1.873
Number of academic jobs	956	0.00	18.00	5.041	2.978
Number of industry jobs	956	0.00	17.00	1.172	1.785
Number of government jobs	956	0.00	7.00	0.431	0.934
Number of medical jobs	956	0.00	6.00	0.066	0.452
Number of consulting jobs	956	0.00	13.00	0.320	1.185
Number of industry jobs/total Jobs	952	0.00	1.00	0.167	0.218
Number of government jobs/total jobs	952	0.00	1.00	0.059	0.124
Total job years (including jobs held concurrently)	956	0.00	183.00	36.052	24.987

Table 4 (cont'd).

Number of job institutions/career length	951	0.02	4.00	.305	.397
Years in academic jobs	956	0.00	160.00	28.394	21.962
Years in industry jobs	956	0.00	41.00	3.892	6.811
Years in government jobs	956	0.00	53.00	2.022	5.696
Years in consulting jobs	956	0.00	48.00	1.514	5.586
Years in medical jobs	956	0.00	20.00	0.230	1.537
Years in industry jobs/total job Years	943	0.00	1.00	0.120	0.197
Years in academic jobs/total job Years	943	0.00	1.00	0.791	0.241
Years in government jobs/total job years	943	0.00	1.00	0.054	0.134
Years in medical jobs/total job Years	943	0.00	0.75	0.006	0.045
Years in consulting jobs/total job years	943	0.00	0.82	0.029	0.093
Ever held industry job? (1=Yes, 0=No)	956	0.00	1.00	0.512	0.500
Ever held government job	956	0.00	1.00	0.264	0.441
First job was industry job? (1=Yes, 0=No)	936	0.00	1.00	0.325	0.469
First job was government job	936	0.00	1.00	0.146	0.354
Career homogeny index (see notes)	937	0.35	72.47	14.470	8.093
Grants					
Number of grants, total [total grants]	381	1.00	130.00	17.688	16.214
Total grants/career length	381	0.03	20.00	1.137	1.313
Number of federal grants	352	1.00	81.00	8.477	8.336
Number of industry grants	193	1.00	74.00	6.021	7.921
Number of grants from other Sources	254	1.00	27.00	4.236	4.182
Federal grants/total grants	352	0.05	1.00	0.530	0.264
Industry grants/total grants	193	0.03	1.00	0.243	0.179

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted.

Table 5. Job Transformations Across and Within Sectors (Frequency and Relative Frequency)

From Job	**To Job**					
Job Sector	Academic	Industry	Government	Consulting	Medical	Totals
Academic	3429	265	158	163	14	4029
Industry	447	249	39	26	2	763
Government	211	53	86	10	1	361
Consulting	150	16	5	105	0	276
Medical	28	3	2	0	28	61
Totals	4265	586	290	304	45	5490

From Job	**To Job**					
Job Sector	Academic	Industry	Government	Consulting	Medical	Totals
Academic	0.85	0.07	0.04	0.04	0.00	1.00
Industry	0.59	0.33	0.05	0.03	0.00	1.00
Government	0.58	0.15	0.24	0.03	0.00	1.00
Consulting	0.54	0.06	0.02	0.38	0.00	1.00
Medical	0.46	0.05	0.03	0.00	0.46	1.00
Totals	3.02	0.65	0.38	0.48	0.47	

Table 6. Tobit Model of Publication Rate (Number of Publications After the Year of the Doctorate as a Proportion of Career Length in Years)

Variable	Regression Coefficient	Standard Error	Asymptotic T-Ratio	P-Value	Normalized Coefficient
Career homogeny index (see notes)	0.049	0.005	3.522	*0.000	0.017
Precocity (see notes)	0.120	0.008	5.270	*0.000	0.042
Held postdoctoral position (1=Yes, 0=No)	-0.309	0.085	-1.265	0.206	-0.107
Triple helix (see notes)	0.277	0.109	0.877	0.380	0.096
First job was industry job? (1=Yes, 0=No)	-0.038	0.097	-0.134	0.894	-0.013
First job was government job	-0.548	0.130	-1.464	0.143	-0.190
Years in industry jobs/total job years	-0.640	0.231	-0.960	0.337	-0.222
Years in government jobs/total job years	1.154	0.328	1.220	0.222	0.400
Number of job institutions/career length	-1.333	0.265	-1.747	0.081	-0.462
Square of above job variable	0.051	0.088	0.200	0.842	0.018
Total grants/career length	0.138	0.043	1.110	0.267	0.048
Industry grants/total grants	0.170	0.304	0.194	0.846	0.059
Federal grants/total grants	0.664	0.130	1.776	0.076	0.230
Doctorate granted before 1972? (1=Yes, 0=No)	1.842	0.135	4.723	*0.000	0.638
Doctorate granted 1972-1980	2.096	0.131	5.559	*0.000	0.727
Doctorate granted 1981-1987	1.537	0.127	4.182	*0.000	0.533
Doctorate granted 1988-1993	0.909	0.109	2.885	*0.004	0.315
Doctorate in biology? (1=Yes, 0=No)	0.013	0.164	0.027	0.979	0.004
Doctorate in computer science	0.543	0.213	0.886	0.376	0.188
Doctorate in engineering	1.037	0.134	2.681	*0.007	0.359
Doctorate in physical sciences	1.380	0.143	3.336	*0.001	0.478
Center 01 (see notes)	-0.801	0.604	-0.460	0.558	-0.278
Center 02	0.086	0.203	0.147	0.989	0.030
Center 04	4.161	0.434	3.323	*0.001	1.442
Center 05	0.600	0.318	0.653	0.566	0.208
Center 06	0.196	0.288	0.236	0.773	0.068
Center 07	-0.439	0.315	-0.483	0.491	-0.152
Center 08	0.211	0.338	0.216	0.875	0.073
Center 09	-2.603	0.399	-2.264	*0.038	-0.902
Center 10	0.954	0.287	1.155	0.249	0.331
Center 11	-0.295	0.664	-0.154	0.760	-0.102
Center 12	1.675	0.276	2.106	*0.024	0.581
Center 13	-1.078	0.538	-0.694	0.569	-0.374
Center 14	0.070	0.289	0.084	0.793	0.024

Table 6 (cont'd).

Center 16	1.085	0.265	1.422	0.143	0.376
Center 18	-0.379	0.365	-0.360	0.731	-0.131
Center 19	0.436	0.363	0.417	0.677	0.151
Center 20	-0.597	0.231	-0.898	0.264	-0.207
Center 21	-0.745	0.313	-0.826	0.338	-0.258
Center 22	-1.223	0.366	-1.159	0.273	-0.424
Center 24	1.318	0.523	0.874	0.337	0.457
Center 25	-2.571	0.535	-1.666	0.174	-0.891
Center 27	1.735	0.435	1.382	0.199	0.601
Center 29	-0.248	0.381	-0.226	0.817	-0.086
Center 31	-0.154	0.381	-0.140	0.797	-0.053
Center 32	1.981	0.408	1.684	0.072	0.687
Center 33	2.799	0.333	2.910	*0.010	0.970
Center 36	0.013	0.269	0.017	0.924	0.005
Center 37	-0.879	0.298	-1.023	0.447	-0.305
Center 38	-0.899	0.277	-1.123	0.246	-0.312
Center 40	-0.617	0.408	-0.525	0.384	-0.214
Center 42	-0.374	0.435	-0.298	0.790	-0.130
Center 43	-1.012	0.408	-0.860	0.458	-0.351
Center 46	0.986	0.669	0.511	0.137	0.342
Center 49	-0.004	0.359	-0.004	0.872	-0.002
Center 50	2.244	0.363	2.145	*0.029	0.778
Center 51	1.329	0.260	1.771	0.081	0.461
Center 53	0.092	0.599	0.053	0.500	0.032
Center 54	-0.259	0.496	-0.181	0.712	-0.090
Center 55	-0.368	0.387	-0.329	0.677	-0.127
Center 56	0.549	0.351	0.542	0.690	0.190
Center 57	0.171	0.391	0.152	0.846	0.059
Center 58	-1.142	0.316	-1.253	0.211	-0.396
Center 59	1.122	0.307	1.268	0.248	0.389
Center 60	0.862	0.220	1.357	0.201	0.299
Center 61	-0.221	0.284	-0.270	0.743	-0.077
Center 62	-0.823	0.374	-0.763	0.491	-0.285
Center 63	0.114	0.368	0.107	0.953	0.040
Center 64	-0.640	0.394	-0.564	0.626	-0.222
Center 66	0.402	0.280	0.498	0.435	0.139
Center 67	0.379	0.225	0.585	0.705	0.131
Center 68	1.728	0.444	1.349	0.137	0.599
Center 69	-0.035	0.259	-0.047	0.914	-0.012
Center 70	-0.125	0.381	-0.113	0.596	-0.043
Center 71	3.459	0.338	3.549	*0.001	1.199
Center 73	-0.314	0.377	-0.289	0.609	-0.109
Center 76	-0.354	0.233	-0.528	0.234	-0.123
Center 81	-1.684	0.418	-1.398	0.213	-0.584
Center 82	-0.029	0.314	-0.032	0.980	-0.010
Center 83	1.376	0.356	1.341	0.164	0.477

Table 6 (cont'd).

Center 84	0.164	0.338	0.169	0.734	0.057
Center 85	1.140	0.232	1.700	0.120	0.395
Center 87	1.555	0.251	2.151	*0.030	0.539
Center 89	0.569	0.273	0.722	0.454	0.197
Center 92	1.192	0.199	2.081	0.074	0.413
CONSTANT	0.038	0.225	0.059	0.100	0.013
ADJPBP		0.008	42.465		0.347

WALD CHI-SQUARE
STATISTIC = 179.37816 WITH
21 D.F. P-VALUE = 0.00000
* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1.

Table 7. Mean Publication and Patent Rate, by Disciplinary Fields

Disciplinary Field	Number of Respondents	Mean Publication Rate	Mean Patent Rate
Agricultural sciences	11	2.913	0.040
Biology			
Physiology/pharmacology	13	4.153	0.024
General biology	50	2.859	0.023
Cell/molecular biology	5	4.082	0.020
Genetics	3	3.522	0.033
Business	9	1.848	0.000
Social sciences/humanities	24	3.063	0.002
Computer sciences	44	3.999	0.121
Engineering			
Aerospace engineering	9	2.088	0.030
Bioengineering	7	4.592	0.252
Chemical engineering	68	4.387	0.195
Civil engineering	38	2.964	0.019
Electrical/Electrical engineering and computer science	153	4.197	0.267
General engineering	9	3.547	0.035
Environmental engineering	9	2.591	0.021
Industrial engineering	23	2.441	0.006
Materials engineering	38	5.135	0.170
Mechanical engineering	54	3.840	0.137
Engineering, other	40	2.756	0.116
Medicine/health sciences	22	5.497	0.041
Mathematical sciences	18	2.714	0.008
Physical sciences			
Physics	101	4.590	0.180
Geosciences	10	1.965	0.000
Astronomy/Astrophysics	6	6.458	0.013
Chemistry	85	4.793	0.278
Physical sciences, other	43	5.099	0.118
Psychology	21	3.946	0.016

Table 8. Tobit Model of Patent Rate (Number of Patents as a Proportion of Career Length in Years)

Variable	Regression Coefficient	Standard Error	Asymptotic T-Ratio	P-Value	Normalized Coefficient
Career homogeny index (see notes)	-0.001	0.006	-0.270	0.787	-0.002
Precocity (see notes)	0.016	0.009	2.917	*0.004	0.027
Held postdoctoral position (1=Yes, 0=No)	-0.056	0.104	-0.900	0.368	-0.094
Triple helix (see notes)	0.055	0.131	0.704	0.482	0.092
First job was industry job? (1=Yes, 0=No)	-0.011	0.115	-0.165	0.869	-0.019
First job was government job	-0.104	0.163	-1.079	0.281	-0.175
Years in industry jobs/total job years	0.828	0.268	5.188	*0.000	1.390
Years in government jobs/total job years	-0.242	0.439	-0.924	0.356	-0.406
Number of job institutions/career length	-0.184	0.350	-0.883	0.378	-0.309
Square of above job variable	0.043	0.118	0.609	0.543	0.072
Total grants/career length	-0.006	0.072	-0.140	0.889	-0.010
Industry grants/total grants	0.415	0.379	1.837	0.066	0.696
Federal grants/total grants	-0.060	0.173	-0.585	0.559	-0.101
Doctorate granted before 1972? (1=Yes, 0=No)	0.306	0.168	3.068	*0.002	0.514
Doctorate granted 1972-1980	0.327	0.163	3.362	*0.001	0.549
Doctorate granted 1981-1987	0.166	0.158	1.757	0.079	0.278
Doctorate granted 1988-1993	0.037	0.143	0.431	0.667	0.061
Doctorate in biology? (1=Yes, 0=No)	0.305	0.230	2.229	*0.026	0.512
Doctorate in computer science	0.457	0.283	2.712	*0.007	0.767
Doctorate in engineering	0.525	0.192	4.581	*0.000	0.881
Doctorate in physical sciences	0.527	0.201	4.405	*0.000	0.885
Center 01 (see notes)	0.288	0.615	0.785	0.424	0.483
Center 02	0.258	0.239	1.808	0.116	0.433
Center 04	1.291	0.446	4.862	*0.000	2.167
Center 05	0.031	0.373	0.138	0.931	0.051
Center 06	0.002	0.340	0.010	0.987	0.003
Center 07	0.080	0.393	0.341	0.739	0.134
Center 08	-0.059	0.390	-0.252	0.762	-0.098
Center 09	-3.719	798.110	-0.008	0.984	-6.241
Center 10	0.188	0.313	1.007	0.315	0.315
Center 11	-0.091	0.702	-0.217	0.688	-0.152
Center 12	0.093	0.314	0.495	0.657	0.156
Center 13	0.357	0.551	1.087	0.294	0.599
Center 14	-0.113	0.353	-0.535	0.553	-0.189
Center 16	0.067	0.296	0.378	0.682	0.112

Table 8 (cont'd).

Center 18	-0.310	0.473	-1.102	0.278	-0.521
Center 19	-0.314	0.438	-1.203	0.208	-0.527
Center 20	-0.609	0.368	-2.775	*0.004	-1.022
Center 21	-3.336	641.970	-0.009	0.983	-5.598
Center 22	-0.258	0.444	-0.972	0.316	-0.432
Center 24	0.075	0.531	0.236	0.814	0.125
Center 25	-0.249	0.563	-0.741	0.493	-0.417
Center 27	0.393	0.456	1.446	0.145	0.659
Center 29	0.169	0.408	0.696	0.513	0.284
Center 31	0.099	0.421	0.393	0.770	0.166
Center 32	0.131	0.460	0.477	0.584	0.219
Center 33	0.190	0.364	0.874	0.377	0.318
Center 36	-0.013	0.316	-0.071	0.876	-0.022
Center 37	0.143	0.364	0.661	0.496	0.240
Center 38	-1.031	0.619	-2.795	*0.005	-1.729
Center 40	-0.067	0.472	-0.239	0.840	-0.113
Center 42	-3.391	943.420	-0.006	0.988	-5.691
Center 43	-0.258	0.491	-0.881	0.349	-0.432
Center 46	-2.728	1336.900	-0.003	0.993	-4.578
Center 49	-0.273	0.633	-0.723	0.464	-0.458
Center 50	0.292	0.386	1.271	0.185	0.490
Center 51	0.526	0.276	3.191	*0.002	0.882
Center 53	-0.339	0.779	-0.730	0.493	-0.569
Center 54	0.021	0.510	0.068	0.986	0.035
Center 55	0.427	0.425	1.686	0.106	0.717
Center 56	-0.261	0.447	-0.979	0.342	-0.438
Center 57	-0.123	0.477	-0.431	0.736	-0.206
Center 58	-0.396	0.425	-1.563	0.103	-0.664
Center 59	0.132	0.328	0.675	0.491	0.221
Center 60	-0.279	0.310	-1.512	0.118	-0.468
Center 61	0.034	0.336	0.168	0.895	0.057
Center 62	-0.604	0.637	-1.591	0.109	-1.013
Center 63	-0.040	0.436	-0.153	0.788	-0.067
Center 64	0.063	0.509	0.208	0.867	0.106
Center 66	0.223	0.304	1.231	0.173	0.374
Center 67	0.042	0.265	0.268	0.870	0.071
Center 68	0.360	0.476	1.270	0.137	0.604
Center 69	-0.426	0.354	-2.022	*0.041	-0.715
Center 70	-0.033	0.448	-0.125	0.813	-0.056
Center 71	-0.274	0.391	-1.177	0.220	-0.460
Center 73	0.150	0.421	0.599	0.518	0.252
Center 76	0.198	0.262	1.271	0.251	0.332
Center 81	-0.194	0.513	-0.634	0.491	-0.325
Center 82	-0.571	0.544	-1.762	0.070	-0.958
Center 83	0.146	0.477	0.515	0.683	0.246

Table 8 (cont'd).

Center 84	-0.072	0.441	-0.273	0.870	-0.121
Center 85	0.043	0.273	0.263	0.738	0.072
Center 87	0.119	0.318	0.626	0.499	0.199
Center 89	-0.201	0.329	-1.024	0.286	-0.337
Center 92	0.138	0.224	1.033	0.316	0.231
CONSTANT	-0.900	0.302	-4.996	*0.000	-1.511
ADJPBP		0.064	26.420		1.678

WALD CHI-SQUARE
STATISTIC = 110.99115 WITH
21 D.F. P-VALUE= 0.00000
* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1.

Table 9. Poisson Model of Patent Rate (Patent Count Divided by Career Length in Years)

Variable	Regression Coefficient	Standard Error	T-Ratio	P-Value	Standardized Coefficient	EXP Coeff
Career homogeny index (see notes)	-0.002	0.013	-0.144	0.886	-0.038	0.998
Precocity (see notes)	0.037	0.015	2.448	*0.014	0.447	1.038
Held postdoctoral position (1=Yes, 0=No)	-0.160	0.228	-0.704	0.481	-0.190	0.852
Triple helix (see notes)	-0.199	0.293	-0.680	0.496	-0.182	0.819
First job was industry job? (1=Yes, 0=No)	-0.093	0.236	-0.395	0.693	-0.112	0.911
First job was government job	-0.167	0.372	-0.449	0.653	-0.152	0.846
Years in industry jobs/total job years	1.759	0.433	4.066	*0.000	0.893	5.808
Years in government jobs/total job years	-0.679	1.110	-0.612	0.541	-0.234	0.507
Number of job institutions/career length	0.259	0.694	0.373	0.709	0.266	1.295
Square of above job variable	-0.066	0.241	-0.273	0.785	-0.188	0.936
Total grants/career length	0.003	0.126	0.023	0.982	0.007	1.003
Industry grants/total grants	1.311	0.593	2.211	*0.027	0.429	3.708
Federal grants/total grants	-0.702	0.442	-1.587	0.113	-0.548	0.496
Doctorate granted before 1972? (1=Yes, 0=No)	0.318	0.363	0.877	0.380	0.326	1.374
Doctorate granted 1972-1980	0.677	0.344	1.970	*0.049	0.682	1.968
Doctorate granted 1981-1987	0.271	0.350	0.775	0.438	0.258	1.312
Doctorate granted 1988-1993	0.133	0.311	0.428	0.669	0.142	1.142
Doctorate in biology? (1=Yes, 0=No)	0.852	0.742	1.148	0.251	0.675	2.344
Doctorate in computer science	1.514	0.736	2.057	*0.040	0.821	4.546
Doctorate in engineering	1.768	0.603	2.933	*0.003	2.277	5.860
Doctorate in physical sciences	1.758	0.608	2.893	*0.004	2.050	5.800
CONSTANT	-4.118	0.715	-5.756	*0.000	0.000	0.016

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1.

Table 10. Differences in Means on Select Variables Between Publication "Stars" and Those Who Are Not Stars

Variable	Pubstar Mean	Nonstar Mean	Pubstar Stdv	Nonstar Stdv	Difference In Means	T-Ratio
Career homogeny index (see notes)	15.32	14.71	7.39	8.22	0.61	0.76
Career length (years since doctorate)	20.44	17.86	11.99	11.42	2.58	*2.02
Doctorate in biological sciences? (1=Yes, 0=No)	0.06	0.12	0.24	0.32	-0.06	*-2.08
Doctorate in computer science	0.03	0.05	0.17	0.21	-0.02	-0.93
Doctorate in engineering	0.45	0.44	0.50	0.50	0.01	0.19
Doctorate in physical sciences	0.41	0.26	0.49	0.44	0.14	*2.75
Doctorate in other field	0.05	0.13	0.22	0.34	-0.08	*-3.16
Total number of publications	236.86	54.83	157.31	54.88	182.03	*11.37
Total number of patents	7.08	0.95	16.42	1.96	6.13	*3.69
Total publications/career length	11.83	2.99	4.25	1.89	8.84	*20.35
Total patents/career length	0.35	0.05	0.75	0.09	0.30	*4.00
Number of job positions, total [total jobs]	7.85	6.61	4.36	3.64	1.24	*2.70
Number of industry jobs	1.18	1.08	1.58	1.72	0.10	0.59
Number of government jobs	0.50	0.44	1.02	0.95	0.06	0.55
Number of industry jobs/total jobs	0.15	0.15	0.17	0.21	-0.01	-0.37
Number of government jobs/total jobs	0.06	0.06	0.11	0.13	0.00	-0.39
Number of job institutions	3.64	3.26	2.21	1.84	0.38	1.64
Total job years (including jobs held concurrently)	43.63	35.34	26.34	25.19	8.29	*2.95
Total job years/career length	3.26	2.39	4.35	1.79	0.87	*1.96
Job institutions/career length	0.37	0.29	0.57	0.35	0.07	1.26
Years in industry jobs	3.89	3.45	6.45	6.42	0.44	0.64
Years in academic jobs	35.47	28.18	22.53	22.14	7.29	*3.03
Years in government jobs	2.59	2.00	7.39	5.61	0.59	0.76
Years in consulting jobs	1.52	1.45	5.84	5.48	0.07	0.11
Years in medical jobs	0.16	0.26	1.01	1.66	-0.10	-0.81
Years in industry jobs/total job years	0.09	0.11	0.15	0.18	-0.02	-0.97
Years in academic jobs/total job years	0.82	0.80	0.19	0.24	0.02	1.02
Years in government jobs/total job years	0.05	0.06	0.13	0.14	0.00	-0.20
Years in medical jobs/total job years	0.01	0.01	0.05	0.05	0.00	-0.03
Years in consulting jobs/total job years	0.02	0.03	0.08	0.09	0.00	-0.33

Table 10 (cont'd).

Ever held industry job? (1=Yes, 0=No)	0.56	0.48	0.50	0.50	0.08	1.50
Ever held government job	0.30	0.27	0.46	0.44	0.03	0.57
Triple helix (1=Yes, 0=No) (see notes)	0.16	0.15	0.37	0.35	0.02	0.42
First job was industry job	0.29	0.30	0.46	0.46	-0.02	-0.31
First job was government job	0.14	0.15	0.35	0.36	-0.01	-0.15
Held postdoctoral position (1=Yes, 0=No)	0.24	0.32	0.43	0.47	-0.07	-1.53
Number of grants, total [total grants]	27.76	16.73	35.59	14.14	11.02	1.66
Total grants/career length	2.10	1.04	3.72	0.81	1.06	1.50
Number of federal grants	12.00	8.26	13.71	8.15	3.74	1.35
Number of industry grants	16.31	5.09	26.97	5.29	11.22	1.66
Number of grants from other sources	5.68	4.15	4.87	4.18	1.53	1.33
Federal grants/total grants	0.52	0.54	0.28	0.27	-0.02	-0.33
Industry grants/total grants	0.28	0.23	0.22	0.18	0.05	0.87
Precocity	7.17	3.17	7.36	4.09	4.01	*5.29
Sex	0.94	0.86	0.24	0.34	0.08	*2.80

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1. (5) Publication star (pubstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of publication adjusted for career length (years since doctorate). Non-stars are defined as those respondents who are not publication or patent stars.

Table 11. Differences in Means on Selected Variables Between Patent "Stars" and Those Who Are Not Stars

Variable	Patstar Mean	Nonstar Mean	Patstar Stdv	Nonstar Stdv	Difference In Means	T-Ratio
Career homogeny index (see notes)	12.30	14.71	7.67	8.22	-2.41	*-2.90
Career length (years since doctorate)	18.40	17.86	11.47	11.42	0.55	0.45
Doctorate in biological sciences? (1=Yes, 0=No)	0.02	0.12	0.14	0.32	-0.10	*-5.31
Doctorate in computer science	0.04	0.05	0.20	0.21	-0.01	-0.38
Doctorate in engineering	0.56	0.44	0.50	0.50	0.12	*2.19
Doctorate in physical sciences	0.36	0.26	0.48	0.44	0.10	1.95
Doctorate in other field	0.02	0.13	0.14	0.34	-0.11	*-5.96
Total number of publications	112.66	54.83	123.34	54.88	57.83	*4.56
Total number of patents	16.73	0.95	20.13	1.96	15.78	*7.80
Total publications/career length	6.08	2.99	5.39	1.89	3.08	*5.58
Total patents/career length	0.94	0.05	0.82	0.09	0.89	*10.82
Number of job positions, total [total jobs]	6.61	6.61	3.81	3.64	0.00	0.00
Number of industry jobs	1.90	1.08	2.20	1.72	0.82	*3.55
Number of government jobs	0.27	0.44	0.81	0.95	-0.17	-1.91
Number of industry jobs/total jobs	0.30	0.15	0.26	0.21	0.15	*5.33
Number of government jobs/total jobs	0.03	0.06	0.08	0.13	-0.03	*-3.41
Number of job institutions	3.19	3.26	1.74	1.84	-0.07	-0.37
Total job years (including jobs held concurrently)	35.89	35.34	22.12	25.19	0.55	0.23
Total job years/career length	2.77	2.39	2.86	1.79	0.38	1.29
Job institutions/career length	0.34	0.29	0.50	0.35	0.05	0.88
Years in industry jobs	7.76	3.45	8.97	6.42	4.31	*4.63
Years in academic jobs	24.40	28.18	18.98	22.14	-3.78	-1.83
Years in government jobs	1.71	2.00	6.13	5.61	-0.30	-0.46
Years in consulting jobs	1.94	1.45	6.13	5.48	0.49	0.76
Years in medical jobs	0.08	0.26	0.71	1.66	-0.18	-1.92
Years in industry jobs/total job years	0.24	0.11	0.27	0.18	0.14	*4.85
Years in academic jobs/total job years	0.67	0.80	0.29	0.24	-0.13	*-4.45
Years in government jobs/total job years	0.04	0.06	0.10	0.14	-0.02	-1.62
Years in medical jobs/total job years	0.00	0.01	0.01	0.05	-0.01	*-3.06
Years in consulting jobs/total job years	0.05	0.03	0.13	0.09	0.02	1.55

Table 11 (cont'd).

Ever held industry job? (1=Yes, 0=No)	0.72	0.48	0.45	0.50	0.24	*4.84
Ever held government job	0.16	0.27	0.37	0.44	-0.11	*-2.63
Triple helix (1=Yes, 0=No) (see notes)	0.10	0.15	0.30	0.35	-0.05	-1.39
First job was industry job	0.52	0.30	0.50	0.46	0.21	*3.95
First job was government job	0.09	0.15	0.29	0.36	-0.06	-1.77
Held postdoctoral position (1=Yes, 0=No)	0.22	0.32	0.42	0.47	-0.09	*-2.08
Number of grants, total [total grants]	30.19	16.73	36.62	14.14	13.45	1.90
Total grants/career length	1.65	1.04	1.55	0.81	0.61	*2.02
Number of federal grants	12.75	8.26	14.57	8.15	4.49	1.49
Number of industry grants	14.55	5.09	24.49	5.29	9.46	1.72
Number of grants from other sources	3.95	4.15	3.81	4.18	-0.20	-0.22
Federal grants/total grants	0.45	0.54	0.21	0.27	-0.09	-1.94
Industry grants/total grants	0.33	0.23	0.18	0.18	0.09	*2.24
Precocity	4.76	3.17	5.70	4.09	1.59	*2.69
Sex	0.97	0.86	0.17	0.34	0.11	*5.04

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1. (5) Patent star (patstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of patents adjusted for career length (years since doctorate). Non-stars are defined as those respondents who are not publication or patent stars.

Table 12. Differences in Means on Selected Variables Between Publication and Patent "Stars"

Variable	Pubstar Mean	Patstar Mean	Pubstar Stdv	Patstar Stdv	Difference In Means	T-Ratio
Career homogeny index (see notes)	15.32	12.30	7.39	7.67	3.02	*2.80
Career length (years since doctorate)	20.44	18.40	11.99	11.47	2.03	1.22
Doctorate in biological sciences? (1=Yes, 0=No)	0.06	0.02	0.24	0.14	0.04	1.46
Doctorate in computer science	0.03	0.04	0.17	0.20	-0.01	-0.37
Doctorate in engineering	0.45	0.56	0.50	0.50	-0.11	-1.50
Doctorate in physical sciences	0.41	0.36	0.49	0.48	0.04	0.64
Doctorate in other field	0.05	0.02	0.22	0.14	0.03	1.16
Total number of publications	236.86	112.66	157.31	123.34	124.20	6.14
Total number of patents	7.08	16.73	16.42	20.13	-9.65	*-3.69
Total publications/career length	11.83	6.08	4.25	5.39	5.76	*8.28
Total patents/career length	0.35	0.94	0.75	0.82	-0.59	*-5.28
Number of job positions, total [total jobs]	7.85	6.61	4.36	3.81	1.24	*2.12
Number of industry jobs	1.18	1.90	1.58	2.20	-0.72	*-2.62
Number of government jobs	0.50	0.27	1.02	0.81	0.23	1.74
Number of industry jobs/total jobs	0.15	0.30	0.17	0.26	-0.15	*-4.86
Number of government jobs/total jobs	0.06	0.03	0.11	0.08	0.03	1.93
Number of job institutions	3.64	3.19	2.21	1.74	0.45	1.59
Total job years (including jobs held concurrently)	43.63	35.89	26.34	22.12	7.74	*2.23
Total job years/career length	3.26	2.77	4.35	2.86	0.49	0.93
Job institutions/career length	0.37	0.34	0.57	0.50	0.03	0.39
Years in industry jobs	3.89	7.76	6.45	8.97	-3.87	*-3.48
Years in academic jobs	35.47	24.40	22.53	18.98	11.07	*3.73
Years in government jobs	2.59	1.71	7.39	6.13	0.88	0.91
Years in consulting jobs	1.52	1.94	5.84	6.13	-0.42	-0.49
Years in medical jobs	0.16	0.08	1.01	0.71	0.08	0.66
Years in industry jobs/total job years	0.09	0.24	0.15	0.27	-0.15	*-4.90
Years in academic jobs/total job years	0.82	0.67	0.19	0.29	0.16	*4.48
Years in government jobs/total job years	0.05	0.04	0.13	0.10	0.02	0.96
Years in medical jobs/total job years	0.01	0.00	0.05	0.01	0.01	0.99
Years in consulting jobs/total job years	0.02	0.05	0.08	0.13	-0.02	-1.53

Table 12 (cont'd).

Ever held industry job? (1=Yes, 0=No)	0.56	0.72	0.50	0.45	-0.16	*-2.30
Ever held government job	0.30	0.16	0.46	0.37	0.13	*2.26
Triple helix (1=Yes, 0=No) (see notes)	0.16	0.10	0.37	0.30	0.06	1.29
First job was industry job	0.29	0.52	0.46	0.50	-0.23	*-3.29
First job was government job	0.14	0.09	0.35	0.29	0.05	1.11
Held postdoctoral position (1=Yes, 0=No)	0.24	0.22	0.43	0.42	0.02	0.37
Number of grants, total [total grants]	27.76	30.19	35.59	36.62	-2.43	-0.25
Total grants/career length	2.10	1.65	3.72	1.55	0.45	0.59
Number of federal grants	12.00	12.75	13.71	14.57	-0.75	-0.19
Number of industry grants	16.31	14.55	26.97	24.49	1.76	0.20
Number of grants from other sources	5.68	3.95	4.87	3.81	1.74	1.23
Federal grants/total grants	0.52	0.45	0.28	0.21	0.07	0.96
Industry grants/total grants	0.28	0.33	0.22	0.18	-0.05	-0.68
Precocity	7.17	4.76	7.36	5.70	2.42	*2.57
Sex	0.94	0.97	0.24	0.17	-0.03	-1.04

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1. (5) Publication star (pubstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of publications adjusted for career length (years since doctorate). Likewise patent star (patstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of patents adjusted for career length (years since doctorate).

Table 13. Differences in Means on Selected Variables Between Publication "Stars" and Those Who are Both Publication and Patent Stars

Variable	Pubstar Mean	Combstar Mean	Pubstar Stdv	Combstar Stdv	Difference In Means	T-Ratio
Career homogeny index (see notes)	15.32	16.79	7.39	8.86	-1.47	-0.76
Career length (years since doctorate)	20.44	19.16	11.99	11.85	1.28	0.48
Doctorate in biological sciences? (1=Yes, 0=No)	0.06	0.00	0.24	0.00	0.06	*2.52
Doctorate in computer science	0.03	0.04	0.17	0.20	-0.01	-0.22
Doctorate in engineering	0.45	0.44	0.50	0.51	0.01	0.08
Doctorate in physical sciences	0.41	0.48	0.49	0.51	-0.07	-0.63
Doctorate in other field	0.05	0.04	0.22	0.20	0.01	0.24
Total number of publications	236.86	229.04	157.31	165.46	7.82	0.21
Total number of patents	7.08	20.32	16.42	28.52	-13.24	*-2.23
Total publications/career length	11.83	12.78	4.25	6.35	-0.95	-0.71
Total patents/career length	0.35	1.08	0.75	1.21	-0.73	*-2.88
Number of job positions, total [total jobs]	7.85	7.44	4.36	4.11	0.41	0.44
Number of industry jobs	1.18	1.32	1.58	1.55	-0.14	-0.39
Number of government jobs	0.50	0.36	1.02	1.25	0.14	0.52
Number of industry jobs/total jobs	0.15	0.19	0.17	0.21	-0.05	-1.02
Number of government jobs/total jobs	0.06	0.02	0.11	0.07	0.03	1.88
Number of job institutions	3.64	3.44	2.21	1.78	0.20	0.48
Total job years (including jobs held concurrently)	43.63	42.80	26.34	29.27	0.83	0.13
Total job years/career length	3.26	2.67	4.35	1.61	0.59	1.07
Job institutions/career length	0.37	0.33	0.57	0.41	0.04	0.37
Years in industry jobs	3.89	5.24	6.45	7.85	-1.35	-0.80
Years in academic jobs	35.47	33.64	22.53	23.87	1.83	0.35
Years in government jobs	2.59	2.44	7.39	10.58	0.15	0.07
Years in consulting jobs	1.52	1.20	5.84	5.60	0.32	0.25
Years in medical jobs	0.16	0.28	1.01	1.40	-0.12	-0.39
Years in industry jobs/total job years	0.09	0.13	0.15	0.19	-0.04	-0.93
Years in academic jobs/total job years	0.82	0.81	0.19	0.23	0.01	0.24
Years in government jobs/total job years	0.05	0.03	0.13	0.09	0.02	1.00
Years in medical jobs/total job years	0.01	0.00	0.05	0.02	0.00	0.54
Years in consulting jobs/total job years	0.02	0.02	0.08	0.10	0.00	0.00

Table 13 (cont'd).

Ever held industry job? (1=Yes, 0=No)	0.56	0.56	0.50	0.51	0.00	0.01
Ever held government job	0.30	0.12	0.46	0.33	0.18	*2.17
Triple helix (1=Yes, 0=No) (see notes)	0.16	0.04	0.37	0.20	0.12	*2.25
First job was industry job	0.29	0.28	0.46	0.46	0.01	0.08
First job was government job	0.14	0.04	0.35	0.20	0.10	1.94
Held postdoctoral position (1=Yes, 0=No)	0.24	0.24	0.43	0.44	0.00	0.05
Number of grants, total [total grants]	27.76	65.00	35.59	59.92	-37.24	-1.47
Total grants/career length	2.10	2.40	3.72	1.72	-0.30	-0.31
Number of federal grants	12.00	24.50	13.71	21.81	-12.50	-1.34
Number of industry grants	16.31	43.50	26.97	45.51	-27.19	-1.15
Number of grants from other sources	5.68	5.50	4.87	6.14	0.18	0.06
Federal grants/total grants	0.52	0.50	0.28	0.28	0.02	0.13
Industry grants/total grants	0.28	0.40	0.22	0.19	-0.12	-1.13
Precocity	7.17	8.40	7.36	8.99	-1.23	-0.63
Sex	0.94	1.00	0.24	0.00	-0.06	*-2.52

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1. (5) Publication star (pubstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of publications adjusted for career length (years since doctorate). Combination star (combstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of publications and patents adjusted for career length (years since doctorate).

Table 14. Differences in Means on Selected Variables Between Patent "Stars" and Those Who are Both Publication and Patent Stars

Variable	Patstar Mean	Combstar Mean	Patstar Stdv	Combstar Stdv	Difference In Means	T-Ratio
Career homogeny index (see notes)	12.30	16.79	7.67	8.86	-4.49	*-2.32
Career length (years since doctorate)	18.40	19.16	11.47	11.85	-0.76	-0.29
Doctorate in biological sciences? (1=Yes, 0=No)	0.02	0.00	0.14	0.00	0.02	1.42
Doctorate in computer science	0.04	0.04	0.20	0.20	0.00	0.01
Doctorate in engineering	0.56	0.44	0.50	0.51	0.12	1.02
Doctorate in physical sciences	0.36	0.48	0.48	0.51	-0.12	-1.03
Doctorate in other field	0.02	0.04	0.14	0.20	-0.02	-0.47
Total number of publications	112.66	229.04	123.34	165.46	-116.38	*-3.29
Total number of patents	16.73	20.32	20.13	28.52	-3.59	-0.59
Total publications/career length	6.08	12.78	5.39	6.35	-6.71	*-4.85
Total patents/career length	0.94	1.08	0.82	1.21	-0.14	-0.56
Number of job positions, total [total jobs]	6.61	7.44	3.81	4.11	-0.83	-0.92
Number of industry jobs	1.90	1.32	2.20	1.55	0.58	1.52
Number of government jobs	0.27	0.36	0.81	1.25	-0.09	-0.33
Number of industry jobs/total jobs	0.30	0.19	0.26	0.21	0.11	*2.12
Number of government jobs/total jobs	0.03	0.02	0.08	0.07	0.01	0.46
Number of job institutions	3.19	3.44	1.74	1.78	-0.25	-0.62
Total job years (including jobs held concurrently)	35.89	42.80	22.12	29.27	-6.91	-1.10
Total job years/career length	2.77	2.67	2.86	1.61	0.10	0.23
Job institutions/career length	0.34	0.33	0.50	0.41	0.01	0.07
Years in industry jobs	7.76	5.24	8.97	7.85	2.52	1.39
Years in academic jobs	24.40	33.64	18.98	23.87	-9.24	-1.80
Years in government jobs	1.71	2.44	6.13	10.58	-0.73	-0.33
Years in consulting jobs	1.94	1.20	6.13	5.60	0.74	0.58
Years in medical jobs	0.08	0.28	0.71	1.40	-0.20	-0.69
Years in industry jobs/total job years	0.24	0.13	0.27	0.19	0.11	*2.45
Years in academic jobs/total job years	0.67	0.81	0.29	0.23	-0.14	*-2.67
Years in government jobs/total job years	0.04	0.03	0.10	0.09	0.01	0.32
Years in medical jobs/total job years	0.00	0.00	0.01	0.02	0.00	-0.64
Years in consulting jobs/total job years	0.05	0.02	0.13	0.10	0.02	0.98

Table 14 (cont'd).

Ever held industry job? (1=Yes, 0=No)	0.72	0.56	0.45	0.51	0.16	1.42
Ever held government job	0.16	0.12	0.37	0.33	0.04	0.55
Triple helix (1=Yes, 0=No) (see notes)	0.10	0.04	0.30	0.20	0.06	1.21
First job was industry job	0.52	0.28	0.50	0.46	0.24	*2.24
First job was government job	0.09	0.04	0.29	0.20	0.05	1.06
Held postdoctoral position (1=Yes, 0=No)	0.22	0.24	0.42	0.44	-0.02	-0.18
Number of grants, total [total grants]	30.19	65.00	36.62	59.92	-34.81	-1.37
Total grants/career length	1.65	2.40	1.55	1.72	-0.75	-0.98
Number of federal grants	12.75	24.50	14.57	21.81	-11.75	-1.25
Number of industry grants	14.55	43.50	24.49	45.51	-28.95	-1.24
Number of grants from other sources	3.95	5.50	3.81	6.14	-1.55	-0.49
Federal grants/total grants	0.45	0.50	0.21	0.28	-0.05	-0.43
Industry grants/total grants	0.33	0.40	0.18	0.19	-0.08	-0.75
Precocity	4.76	8.40	5.70	8.99	-3.64	-1.93
Sex	0.97	1.00	0.17	0.00	-0.03	-1.75

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1. (5) Patent star (patstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of patents adjusted for career length (years since doctorate). Combination star (combstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of publications and patents adjusted for career length (years since doctorate).

Table 15. Differences in Means on Selected Variables Between Those Who are Both Publication and Patent "Stars" and those who are not stars.

Variable	Combstar Mean	Nonstar Mean	Combstar Stdv	Nonstar Stdv	Difference In Means	T-Ratio
Career homogeny index (see notes)	16.79	14.71	7.67	8.22	2.08	1.16
Career length (years since doctorate)	19.16	17.86	11.47	11.42	1.30	0.54
Doctorate in biological sciences? (1=Yes, 0=No)	0.00	0.12	0.14	0.32	-0.12	*-10.20
Doctorate in computer science	0.04	0.05	0.20	0.21	-0.01	-0.21
Doctorate in engineering	0.44	0.44	0.50	0.50	0.00	0.01
Doctorate in physical sciences	0.48	0.26	0.48	0.44	0.22	*2.09
Doctorate in other field	0.04	0.13	0.14	0.34	-0.09	*-2.19
Total number of publications	229.04	54.83	123.34	54.88	174.21	*5.26
Total number of patents	20.32	0.95	20.13	1.96	19.37	*3.40
Total publications/career length	12.78	2.99	5.39	1.89	9.79	*7.69
Total patents/career length	1.08	0.05	0.82	0.09	1.04	*4.26
Number of job positions, total [total jobs]	7.44	6.61	3.81	3.64	0.83	1.00
Number of industry jobs	1.32	1.08	2.20	1.72	0.24	0.75
Number of government jobs	0.36	0.44	0.81	0.95	-0.08	-0.32
Number of industry jobs/total jobs	0.19	0.15	0.26	0.21	0.04	0.92
Number of government jobs/total jobs	0.02	0.06	0.08	0.13	-0.04	*-2.59
Number of job institutions	3.44	3.26	1.74	1.84	0.18	0.49
Total job years (including jobs held concurrently)	42.80	35.34	22.12	25.19	7.46	1.26
Total job years/career length	2.67	2.39	2.86	1.79	0.28	0.86
Job institutions/career length	0.33	0.29	0.50	0.35	0.04	0.46
Years in industry jobs	5.24	3.45	8.97	6.42	1.79	1.13
Years in academic jobs	33.64	28.18	18.98	22.14	5.46	1.13
Years in government jobs	2.44	2.00	6.13	5.61	0.44	0.21
Years in consulting jobs	1.20	1.45	6.13	5.48	-0.25	-0.22
Years in medical jobs	0.28	0.26	0.71	1.66	0.02	0.07
Years in industry jobs/total job years	0.13	0.11	0.27	0.18	0.02	0.58
Years in academic jobs/total job years	0.81	0.80	0.29	0.24	0.01	0.21
Years in government jobs/total job years	0.03	0.06	0.10	0.14	-0.03	-1.33
Years in medical jobs/total job years	0.00	0.01	0.01	0.05	0.00	-1.01
Years in consulting jobs/total job years	0.02	0.03	0.13	0.09	0.00	-0.14

Table 15 (cont'd).

Ever held industry job? (1=Yes, 0=No)	0.56	0.48	0.45	0.50	0.08	0.77
Ever held government job	0.12	0.27	0.37	0.44	-0.15	*-2.17
Triple helix (1=Yes, 0=No) (see notes)	0.04	0.15	0.30	0.35	-0.11	*-2.54
First job was industry job	0.28	0.30	0.50	0.46	-0.02	-0.26
First job was government job	0.04	0.15	0.29	0.36	-0.11	*-2.62
Held postdoctoral position (1=Yes, 0=No)	0.24	0.32	0.42	0.47	-0.08	-0.86
Number of grants, total [total grants]	65.00	16.73	36.62	14.14	48.27	*1.97
Total grants/career length	2.40	1.04	1.55	0.81	1.36	1.93
Number of federal grants	24.50	8.26	14.57	8.15	16.24	1.82
Number of industry grants	43.50	5.09	24.49	5.29	38.41	1.69
Number of grants from other sources	5.50	4.15	3.81	4.18	1.35	0.44
Federal grants/total grants	0.50	0.54	0.21	0.27	-0.04	-0.31
Industry grants/total grants	0.40	0.23	0.18	0.18	0.17	1.81
Precocity	8.40	3.17	5.70	4.09	5.23	*2.90
Sex	1.00	0.86	0.17	0.34	0.14	*11.13

* Indicates P-Value < .05

Notes: (1) Homogeny is an index of career patterns over time. It is the sum of conditional job probabilities of moving from one job to another corrected for total number of jobs. A high value means the career path is tending toward a common one. A low value means that few others in the data set made such job transformations. (2) Precocity is the number of publications during or before the year the doctorate was granted. (3) Triple Helix is a dummy variable taking on the value of 1 when the respondent has had at least one job in all three sectors (academia, industry, and government) otherwise the value is 0. (4) Center variables are dummy variables taking on a value of 1 when the researchers is affiliated with that particular NSF, DOE, or DOD center otherwise the value is 0. For a list of center names, see Table 1. (5) Combination star (combstar) is defined to be respondents who are in the top ten percent of the distribution of the total number of publications and patents adjusted for career length (years since doctorate). Non-stars are defined as those respondents who are not publication or patent stars.

Table 16. Mean Number of Publications For Five Years Before and After a Job Transformation from and to Industry (Standard Deviation, 95% Confidence Intervals)

From Industry

	Mean	Std. Dev	Lower bound CI	Upper bound CI
Mean before	1.50	2.84	1.23	1.77
Mean after	2.60	3.55	2.27	2.93

To Industry

	Mean	Std. Dev	Lower bound CI	Upper bound CI
Mean before	1.81	3.01	1.44	2.18
Mean after	2.57	3.12	2.19	2.95

Note: CI stands for confidence interval.

Table 17. Neural Network Models, Estimated Output Values for Selected Input Values

Variable	Variable Value	Publication Rate Estimate	Patent Rate Estimate
Industry grants/total grants	0.00	20.204	0.164
	0.05	20.058	0.171
	0.10	19.827	0.179
	0.15	19.464	0.188
	0.20	18.908	0.197
	0.25	18.095	0.208
	0.30	16.990	0.220
	0.35	15.643	0.233
	0.40	14.218	0.247
	0.45	12.938	0.262
	0.50	11.969	0.280
	0.55	11.372	0.299
	0.60	11.134	0.320
	0.65	11.217	0.344
	0.70	11.593	0.370
	0.75	12.253	0.398
	0.80	13.201	0.430
	0.85	14.440	0.465
	0.90	15.948	0.503
	0.95	17.654	0.545
	1.00	19.429	0.591
Federal grants/total grants	0.00	9.010	0.235
	0.05	9.219	0.238
	0.10	9.442	0.242
	0.15	9.679	0.246
	0.20	9.933	0.250
	0.25	10.206	0.254
	0.30	10.502	0.258
	0.35	10.822	0.263
	0.40	11.171	0.269
	0.45	11.552	0.274
	0.50	11.969	0.280
	0.55	12.426	0.286
	0.60	12.929	0.293
	0.65	13.481	0.299
	0.70	14.087	0.307
	0.75	14.750	0.314
	0.80	15.472	0.322
	0.85	16.251	0.331
	0.90	17.085	0.340
	0.95	17.965	0.349
	1.00	18.876	0.359

Table 17 (cont'd).

Total grants/career length	0	11.786	0.275
	5	11.945	0.278
	10	11.969	0.280
	15	11.808	0.281
	20	11.403	0.283
Years in government jobs/total job years	0.00	21.749	0.276
	0.05	21.213	0.277
	0.10	20.563	0.277
	0.15	19.777	0.278
	0.20	18.845	0.278
	0.25	17.773	0.278
	0.30	16.596	0.279
	0.35	15.367	0.279
	0.40	14.151	0.279
	0.45	13.006	0.280
	0.50	11.969	0.280
	0.55	11.054	0.280
	0.60	10.262	0.280
	0.65	9.587	0.280
	0.70	9.021	0.281
	0.75	8.566	0.281
	0.80	8.232	0.281
	0.85	8.033	0.281
	0.90	7.987	0.281
	0.95	8.100	0.281
	1.00	8.361	0.281
Years in industry jobs/total job years	0.00	23.934	0.130
	0.05	22.986	0.139
	0.10	21.806	0.148
	0.15	20.445	0.158
	0.20	18.981	0.170
	0.25	17.503	0.183
	0.30	16.092	0.197
	0.35	14.809	0.214
	0.40	13.689	0.233
	0.45	12.744	0.255
	0.50	11.969	0.280
	0.55	11.345	0.309
	0.60	10.850	0.342
	0.65	10.458	0.380
	0.70	10.145	0.424
	0.75	9.888	0.474
	0.80	9.669	0.531
	0.85	9.471	0.596
	0.90	9.280	0.669
	0.95	9.086	0.750
	1.00	8.876	0.840

Table 17 (cont'd).

Total job institutions/career length	0.0	20.046	0.587
	0.1	19.763	0.564
	0.2	19.461	0.541
	0.3	19.141	0.520
	0.4	18.801	0.499
	0.5	18.443	0.480
	0.6	18.067	0.461
	0.7	17.674	0.444
	0.8	17.267	0.427
	0.9	16.847	0.411
	1.0	16.417	0.395
	1.1	15.978	0.381
	1.2	15.534	0.367
	1.3	15.086	0.354
	1.4	14.636	0.342
	1.5	14.187	0.330
	1.6	13.739	0.319
	1.7	13.293	0.308
	1.8	12.850	0.298
	1.9	12.408	0.289
	2.0	11.969	0.280
	2.1	11.529	0.271
	2.2	11.089	0.263
	2.3	10.646	0.256
	2.4	10.199	0.248
	2.5	9.746	0.241
	2.6	9.287	0.235
	2.7	8.820	0.229
	2.8	8.347	0.223
	2.9	7.866	0.217
	3.0	7.381	0.212
	3.1	6.893	0.207
	3.2	6.405	0.202
	3.3	5.923	0.198
	3.4	5.449	0.194
	3.5	4.989	0.190
	3.6	4.549	0.186
	3.7	4.131	0.182
	3.8	3.739	0.179
	3.9	3.377	0.175
	4.0	3.045	0.172

Table 17 (cont'd).

Career homogeny index	0	8.902	0.155
	5	9.008	0.158
	10	9.129	0.163
	15	9.273	0.171
	20	9.471	0.184
	25	9.799	0.202
	30	10.410	0.229
	35	11.566	0.268
	40	13.502	0.323
	45	15.983	0.400
	50	18.178	0.507
	55	19.565	0.651
	60	20.271	0.838
	65	20.587	1.067
	72	20.729	1.473
Center 87	Yes	6.312	0.461
	No	11.333	0.196
Center 71	Yes	14.038	0.299
	No	9.843	0.264
Center 69	Yes	10.281	0.245
	No	17.526	0.323
Center 51	Yes	9.204	0.246
	No	13.998	0.324
Center 50	Yes	14.492	0.431
	No	13.798	0.222
Center 38	Yes	13.397	0.175
	No	11.052	0.468
Center 33	Yes	19.135	0.314
	No	8.625	0.260
Center 20	Yes	10.744	0.221
	No	12.740	0.356
Center 12	Yes	11.754	0.333
	No	13.278	0.237
Center 04	Yes	20.555	1.398
	No	10.345	0.096
Precocity	0	1.596	0.180
	5	1.463	0.196
	10	1.576	0.216
	15	3.905	0.240
	20	10.797	0.270
	25	13.396	0.306
	30	15.263	0.352
	35	18.606	0.407
	40	22.635	0.475
	43	24.324	0.523

Table 17 (cont'd).

First job was government job	Yes	4.808	0.184
	No	18.949	0.493
First job was industry job	Yes	8.775	0.294
	No	17.034	0.258
Triple Helix	Yes	18.161	0.208
	No	9.756	0.392
Held postdoctoral position	Yes	9.801	0.139
	No	20.566	0.766
Held government job	Yes	13.502	0.224
	No	12.011	0.395
Held industry job	Yes	11.173	0.829
	No	15.285	0.161
Doctorate in physical sciences	Yes	20.276	0.909
	No	9.411	0.207
Doctorate in engineering	Yes	12.772	0.242
	No	11.960	0.555
Doctorate in computer science	Yes	8.923	0.214
	No	17.130	0.380
Doctorate in biological science	Yes	10.865	0.154
	No	18.167	0.590
Doctorate granted 1988-1993	Yes	6.218	0.282
	No	11.878	0.239
Doctorate granted 1981-1987	Yes	12.492	0.181
	No	12.051	0.513
Doctorate granted 1972-1980	Yes	15.693	0.831
	No	5.848	0.151
Doctorate granted before 1972	Yes	14.993	0.239
	No	14.556	0.333

Table 18. Summary of Tobit, Poisson, and Neural Network Models—Relationship Between Selected Independent Variables (Inputs) and Dependent Variables (Outputs)

Variable	Tobit Model--Publication Rate	Tobit Model--Patent Rate	Poisson Model--Patent Rate	Neural Network Model--Publication Rate	Neural Network Model--Patent Rate
Career homogeny index	0*	0	0	0>+	0>+
Precocity	0*	0*	0*	0>+	0>+
Held postdoctoral position (1=Yes, 0=No)	-	-	-	-	-
Triple Helix	+	0	-	+	-
First job was industry job? (1=Yes, 0=No)	0	-	-	-	0
First job was government job	-	-	-	-	-
Years in industry jobs/total job years	-	+*	+*	-	+
Years in government jobs/total job years	+	-	-	-	0
Total grants/career length	0	0	0	0	0
Industry grants/total grants	0	+	+*	-	+
Federal grants/total grants	+	+	-	+	+
Number of job institutions/career length	-	-	+	-	-
Doctorate granted before 1972? (1=Yes, 0=No)	+*	+*	+	0	0
Doctorate granted 1972-1980	+*	+*	+*	+	+
Doctorate granted 1981-1987	+*	+	+	+	-
Doctorate granted 1988-1993	+*	0	+	-	0
Doctorate in biological science (1=Yes, 0=No)	0	+*	+	-	-
Doctorate in computer science	+	+*	+*	-	-
Doctorate in engineering	+*	+*	+*	+	+
Doctorate in physical sciences	+*	+*	+*	+	+

Notes: + Indicates variable had positive relationship with dependent variable, - indicates a negative relationship, and 0 indicates no relationship of substantive significance. For the neural models judgment was made based on output queries for inputs of various levels). 0>+ Indicates that the relationship was manifest only for extreme positive values of the input. * Indicates statistical significance at the level of .05 or below (not applicable to neural models).

Table 19. Summary of Differences in Means in Selected Variables Between Publication "Stars," Patent "Stars," Those Who are Both Publication and Patent Stars, and Those Who are Not Stars (Nonstars)

	Means			
Variable	Pubstar	Patstar	Combstar	Nonstar
Career homogeny index	*15.32	§12.30	+16.79	14.71
Precocity	*§7.17	§4.76	§8.40	3.17
Held postdoctoral position (1=Yes, 0=No)	0.24	§0.22	0.24	0.32
Triple helix	0.16	0.10	o§0.04	0.15
First job was industry job? (1=Yes, 0=No)	*0.29	§0.52	+0.28	0.30
First job was government job	0.14	0.09	§0.04	0.15
Held industry job (1=Yes, 0=No)	*0.56	§0.72	0.56	0.48
Held government job	*0.30	§0.16	o§0.12	0.27
Years in industry jobs/total job years	*0.09	§0.24	+0.13	0.11
Years in government jobs/total job years	0.05	0.04	0.03	0.06
Job institutions/career length	0.37	0.34	0.33	0.29
Total grants/career length	2.10	1.65	2.40	1.04
Industry grants/total grants	0.28	§0.33	0.40	0.23
Federal grants/total grants	0.52	0.45	0.50	0.54
Doctorate in biological science (1=Yes, 0=No)	§0.06	§0.02	o§0.00	0.12
Doctorate in computer science	0.03	0.04	0.04	0.05
Doctorate in engineering	0.45	§0.56	0.44	0.44
Doctorate in physical sciences	§0.41	0.36	§0.48	0.26

Notes: * Indicates a statistically significant difference in means between publication star and patent star. + Indicates a statistically significant difference between patent star and combination star. o Indicates a statistically significant difference between publication star and combination star. § Indicates a statistically significant difference in means with non-star.

APPENDIX B: Tobit Model Diagnostics

One problem with limited dependent variable models, including the Tobit model, is that they are sensitive to model misspecification, heteroscedasticity, and non-normal error distributions (Kennedy, 1998). It is likely the models present here will suffer from one or more of these problems due to the highly skewed nature of productivity and appropriate diagnostics are discussed below. In the presence of non-constant variance of the errors terms (heteroscedasticity), maximum likelihood estimators remain consistent although they are no longer efficient. As a result, confidence intervals and P-values for estimators may be larger than should be the case.

To test for heteroscedasticity residual plots were examined and a Modified- Levene test was constructed. As shown in Figures 20 and 21 the residuals plotted against the predicted dependent variables appear to exhibit non-constant variance due to the presence of outliers. It also appears, not surprisingly, that there is possibly increasing variance as the value of the predicted dependent variable increases. However, this is difficult to judge given the distorting effects of the outliers on the Y axis of the graphs.

In Figures 22 and 23 outliers above three standard deviations from the mean have been removed in order to examine more closely the relationship between the residuals and the dependent variable for the majority of the cases. There does seem to be a V or funnel shape to the distribution, which again indicates the presence of heteroscedasticity. However, there does not appear to be a relationship[45] other than increasing variance, so it may be the case that important variables have not been omitted from the models.

[45] Other scatter plots were run on the independent variables that are not shown here.

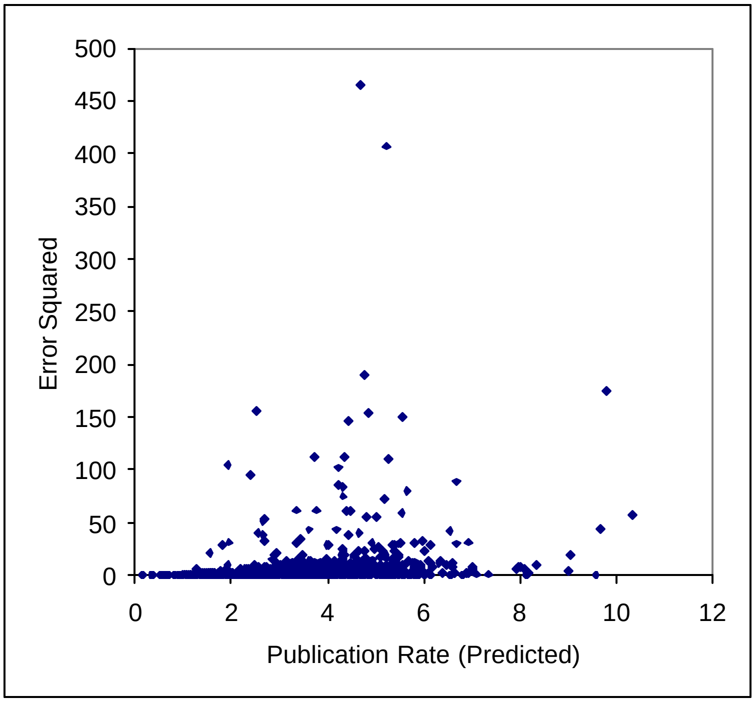

Figure 20. Scatter Plot of Predicted Publication Rate by Error Squared (all cases)

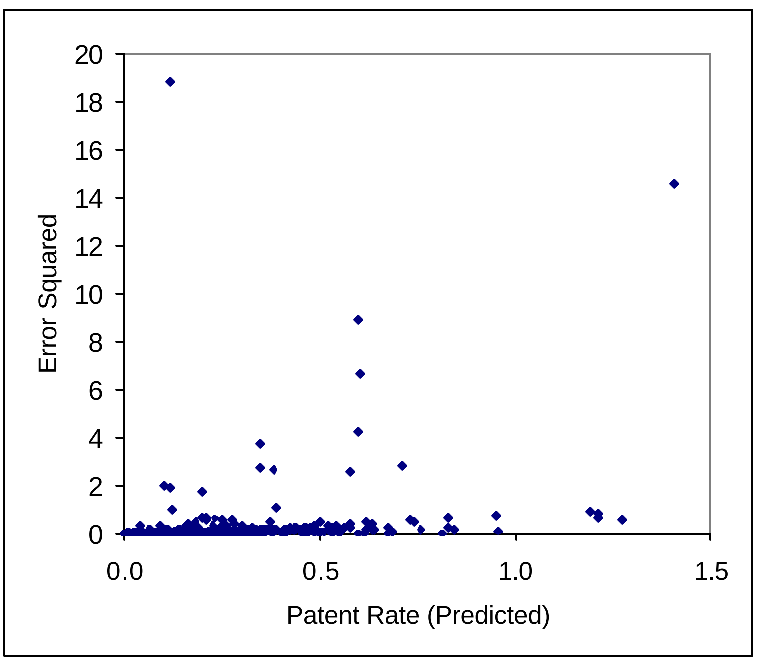

Figure 21. Scatter Plot of Predicted Patent Rate by Error Squared (all cases)

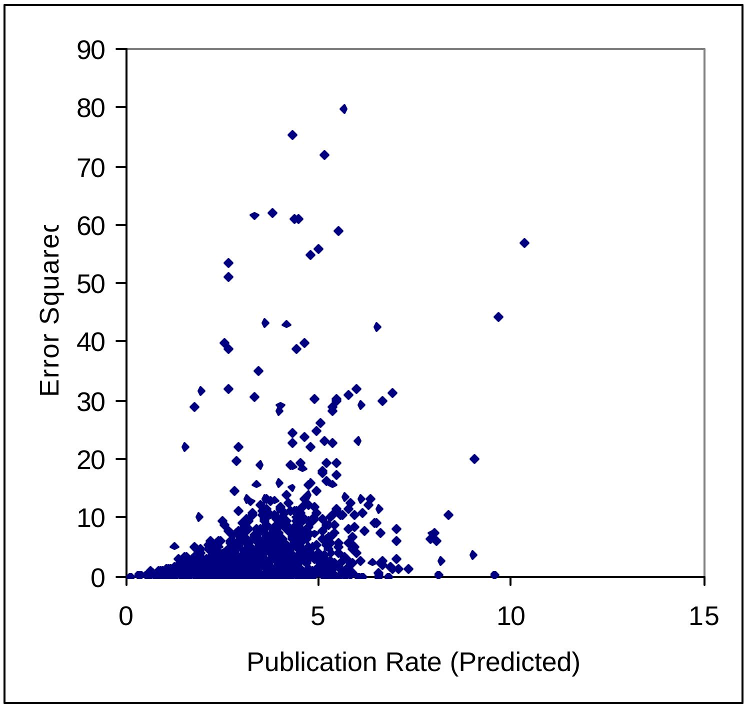

Figure 22. Scatter Plot of Predicted Publication Rate by Error Squared (outliers above three standard deviations of the mean removed)

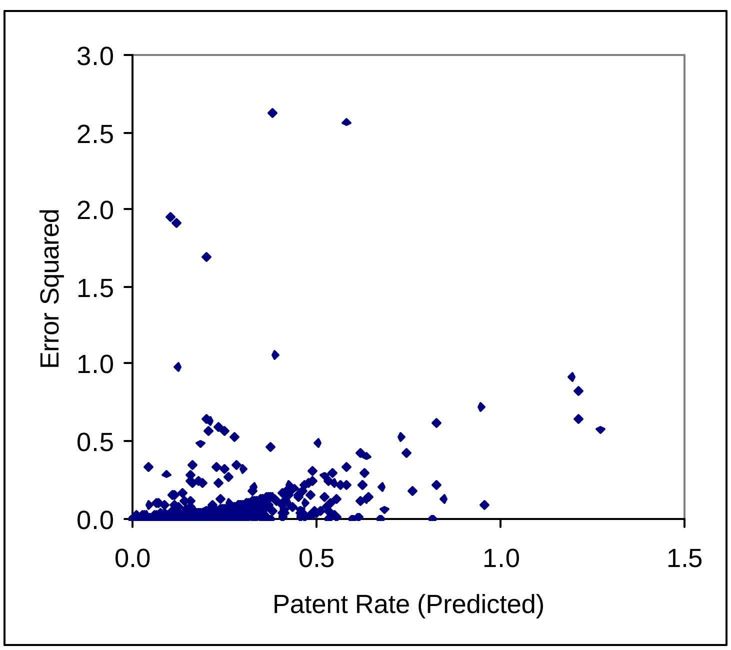

Figure 23. Scatter Plot of Predicted Publication Rate by Error Squared (outliers above three standard deviations of the mean removed)

For this reason, a Modified Levene test was run on each model. The results[46] confirm that there is increasing variance in the errors as the predicted value of publication and patent rates increase.

Departures from the normality assumption are often thought to be less critical than the presence of heteroscedasticity because point estimation remains unbiased and, as sample size increases, estimators will generally tend to be distributed normally despite the violation of the normality assumption (Gujarati, 1995). Nonetheless, as shown in Figures 24 and 25, the error terms appear to be distributed non-normally[47]. The error term distribution on the publication rate model had a skewness statistic of 2.34 where normal distributions hover around 0 and where a positive skewness value indicates a distribution with a long right tail (this also confirms heteroscedasticity). The kurtosis value of 9.94 indicates that the distribution of errors cluster around the central point more than is the case with the normal distribution and has longer tails. The distribution of error terms for patent rate was even more skewed (5.90) and kurtotic (55.79).

---

[46] The Modified Levene statistic for publication rate was –5.88 and –6.49 for patent rate.

[47] One possible remedy to the presence of heteroscedasticity and non-normality is to transform the dependent variable (and possibly the independent variables). However, this complicates model interpretation and is unlikely to restore homoscedasticity when the dependent variable contains large numbers of zero values.

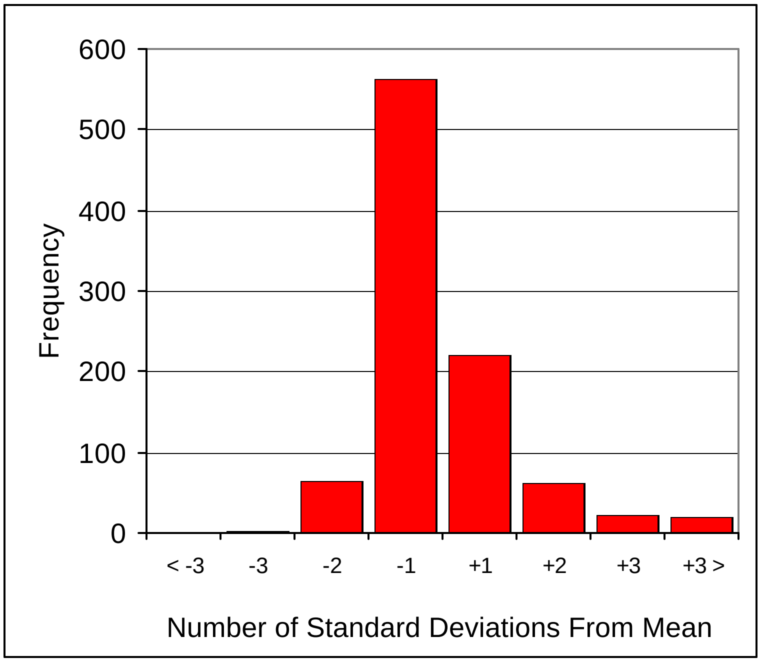

Figure 24. Publication Rate Normality Diagnostic—Number of Observations 1, 2, 3, and More Than 3 Standard Deviations Above or Below the Mean Error

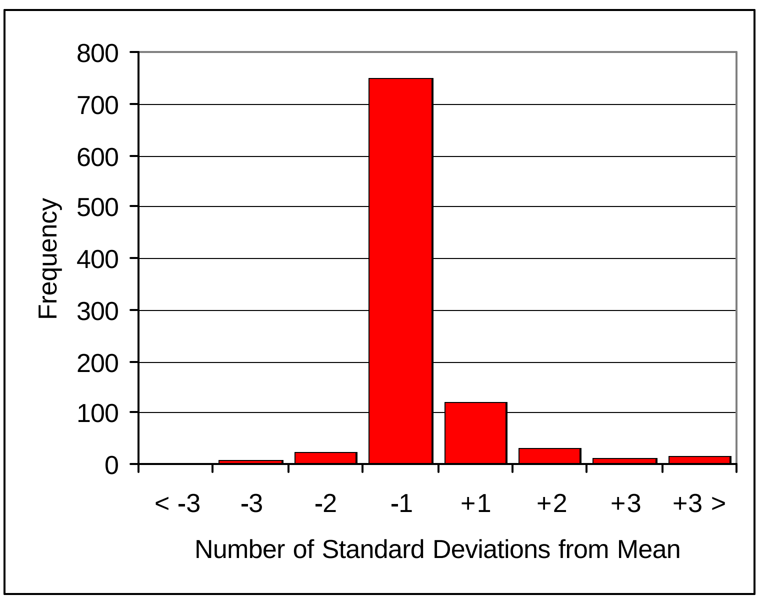

Figure 25. Patent Rate Normality Diagnostic—Number of Observations 1, 2, 3, and More Than 3 Standard Deviations Above or Below the Mean Error

Finally, to test for multicollinearity the Variance Inflation Factors for each of the variables was examined. As a rule of thumb factors above 10 indicate collinearity; high factors in general suggest that standard errors of the estimates will be inflated. For most variables in these models the factors were quite low (e.g., below 3.0). However, for engineering (8.6), physical science (6.5), and computer science (5.1) the factors are approaching the high levels, indicating that these variables are at least moderately correlated with other variables in the model.

REFERENCES

Allen, T. J., and R. Katz 1986.The dual ladder: Motivational solution or managerial delusion? *R&D Management* 16(2): 185-197.

Allen, T. J., and R. Katz 1992. Age, education and the technical ladder. *IEEE Transactions on Engineering Management* 39: 237-245.

Allison, P. D., and J. S. Long 1987. Inter-university mobility of academic scientists. *American Sociological Review* 52: 643-652.

Allison, P. D., and J. S. Long 1990. Departmental effects on scientific productivity. *American Sociological Review* 55: 469-478.

Allison, P. D., J. S. Long. and T. K. Krauze 1982. Cumulative advantage and inequality in science. *American Sociological Review* 47: 615-25.

Allison, P. D., and J. A. Stewart 1974. Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review* 39: 596-606.

Arminger, G. 1995. Specification and Estimation of Mean Structures: Regression Models. In *Handbook of statistical modeling for the social and behavioral sciences*. G. Arminger, C. Clogg, and M. Sobel eds.

Bartel, A. P., and N. Sicherman, 1993. Technological-change and retirement decisions of older workers. *Journal of Labor Economics* 11, 162-183.

Bayer, A. E., and J. Folger 1966. Some correlates of a citation measure of productivity in science. *Sociology of Education* 39: 381-90.

Bayer, A. E., and J. C. Smart 1991. Career publication patterns and collaborative 'styles' in American academic science. *Journal of Higher Education* 62: 613-636.

Becker, G. 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economics* 70: S9-S49.

Becker, G. 1964. *Human capital: A theoretical and empirical analysis, with a special reference to education*. Chicago: University of Chicago Press.

Bengio, Y. 1999. Markovian models for sequential data. *Neural Computing Surveys* 2: 129-162. http://www.icsi.berkeley.edu/jagota/NCS.

Bonzi, S. 1992. Trends in research productivity among senior faculty. *Information Processing & Management* 28(1): 111-120.

Bourdieu, P. 1986. The forms of social capital. In *Handbook of theory and research for the sociology of education*. J. G. Richardson ed., New York: Greenwood.

Bourdieu, P., and L. Wacquant 1992. *An invitation to reflexive sociology*. Chicago: University of Chicago Press.

Bowden, V. 1997. The career states system model: A new approach to analysing careers. *British Journal of Guidance and Counseling* 25(4): 473-491.

Bozeman, B., J. S. Dietz, and M. Gaughan 2001. Scientific and technical human capital: An alternative approach to R&D evaluation. *International Journal of Technology Management* 22(8): 716-740.

Bozeman B. and J. Melkers 1993. *Evaluating R&D impacts: Methods and practice*. Boston: Kluwer Academic Publishers.

Bozeman, B. and J. Rogers 2000. Strategic management of government-sponsored R&D portfolios: Project outputs and scientific and technical human capital. *Environment and Planning C: Government and Policy*.

Bozeman, B., and J. D. Rogers 2002. A churn model of scientific knowledge value: Internet researchers as a knowledge value collective. *Research Policy* 31: 769-794.

Bozeman, B., J. Rogers, D. Roessner, H. Klein, and J. Park 1998. The R&D value mapping project: Final report. *Report to the Department of Energy, Office of Basic Energy Sciences*. Atlanta, GA: Georgia Institute of Technology.

Bozeman, B., J. Rogers, D. Roessner, H. Klein, J. Park, G. Kingsley, S. Silver, S. Sreedhar, T. Sykes, M. Kelman, and J, Dietz 1999. The research value mapping project: Qualitative-quantitative case studies of research projects funded by the Office of Basic Energy Sciences. *Final report to the Department of Energy, Office of Basic Energy Sciences*. Atlanta, GA: Georgia Institute of Technology.

Bozeman, B., J. D. Roessner, J. Rogers, and H. Klein 1997. The R&D value mapping Project: Annual report, 1997. *Report to the Department of Energy, Office of Basic Energy Sciences*. Atlanta, GA: Georgia Institute of Technology.

Brancomb, L., and J. Keller, eds.1998. *Investing in innovation: Creating a research and innovation policy that works*. Cambridge, MA: MIT Press.

Burns, T. 1994. *The management of innovation*. New York: Oxford University Press.

Burt, R. S. 1992. *Structural holes: The social structure of competition*. Cambridge: Harvard University Press.

Burt, R. S. 1997a. A note on social capital and network content. *Social Networks* 19: 355-373.

Burt, R. S. 1997b. The contingent value of social capital. *Administrative Science Quarterly* 42: 339-365.

Callon, M., and J. Law, J. 1989. On construction of sociotechnical networks: Content and context revisited. *Knowledge and Society: Studies in the Sociology of Science Past and Present* 8: 57-85.

Callon, M. 1992. "The dynamics of techno-economic networks," In Coombs, R., P. Saviotti, and V. Walsh, eds., *Technological change and company strategies*. London: Academic Press. 72-102.

Callon, M., J. P. Courtial, P. Crance, P. Laredo, P. Mauguin, V. Rabeharisoa, Y.A. Rocher, and D. Vinck 1991. Tools for the evaluation of technological programmes: An account of work done at the Centre for the Sociology of Innovation. *Technology Analysis & Strategic Management* 3: 3-41.

Caracostas, P., and U. Muldur 1998. *Society, The Endless Frontier: A European Vision of Research and Innovation Policies for the 21st Century*. Italy: Office for Official Publications of the European Communities, European Commission.

Carley, K. M., N. P. Hummon, and M. Harty 1993. Scientific influence--An analysis of the main path structure in The Journal of Conflict-Resolution. *Knowledge Creation Diffusion Utilization* 14: 417-447.

Clemente, F. 1973. Early career determinants of research productivity. *American Journal of Sociology* 79: 409-419.

Cole, J. R. 1970. Patterns of intellectual influence in scientific research. *Sociology of Education* 43: 377-403.

Cole, J. R., and S. Cole 1971. Measuring the quality of sociological research: Problems in the use of the Science Citation Index. *The American Sociologist* 6: 23-29.

Cole, J. R., and S. Cole 1973. *Social Stratification in Science*. Chicago: Chicago University Press.

Cole, S. 1979. Age and scientific performance. *American Journal of Sociology* 84: 958-977.

Cole, S., and J. R. Cole 1967. Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review* 32: 377-390.

Coleman, J. C. 1988. Social capital in the creation human capital. *American Journal of Sociology* 94: S95-S120.

Coleman, J. C. 1990. *Foundations of Social Theory*. Cambridge: The Belknap Press of Harvard University Press.

Coleman, J. C. 1993. The impact of Becker, Gary on sociology. *Acta Sociologica*, 36, 169-178.

Constant, D., L. Sproull, and S, Kiesler 1996. The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science* 7: 119-135.

Courtial, J. P., T. Chlik, and M. Callon 1994. A model for social interaction between cognition and action through a key-word simulation of knowledge growth. *Scientometrics* 31: 173-192.

Cozzens, S. 1989. What do citations count? The rhetoric-first model. *Scientometrics* 15: 437-447.

Cozzens, S., S. Popper, J. Bonomo, K. Koizumi, and A. Flanagan 1994. Methods for evaluating fundamental science. *Report Prepared for the Office of Science and Technology Policy*. Washington, D.C.: RAND Corporation, Critical Technologies Institute.

Crane, D. 1965. Scientists at major and minor universities: A study of productivity and recognition. *American Sociological Review* 30: 699-714.

Crane, D. 1969. Social structure in a group of scientists: A test of the invisible college hypothesis. *American Sociological Review* 34: 335-352.

Crane, D. 1970. The academic marketplace revisited: A study of faculty mobility using the Cartter ratings. *American Journal of Sociology* 75: 953-964.

Crane, D. 1972. *The invisible college*. Chicago: University of Chicago Press.

Crittenden, K., and R. Hill 1971. Coding reliability and validity of interview data. *American Sociological Review* 36(6): 1073-1080.

Dalton, G. W., and P. H. Thompson 1971. Accelerating obsolescence of older engineers. *Harvard Business Review* 49: 57-67.

Debackere, K., D. Buyens, and T. Vandenbossche 1997.Strategic career development for R&D professionals: Lessons from field research. *Technovation* 17(2): 53-62.

Debackere, K., and M. A. Rappa 1994. Technological communities and the diffusion of knowledge—A replication and validation. *R &D Management* 24(4): 355-371.
Diamond, A. M. Jr. 1984. An economic model of the life-cycle research productivity of scientists. *Scientometrics* 6(3): 189-196.

Diamond, A. M., Jr. 1986. The life cycle research productivity of mathematicians and scientists. *Journal of Gerontology* 41: 520-525.

Dietz, J. S. 2000. Building a social capital model of research development: The case of the Experimental Program to Stimulate Competitive Research. *Science and Public Policy*, 27: 137-145.

Dietz, J. S., I. Chompalov, B. Bozeman, E. O. Lane, and J. Park, J. 2000. Using the curriculum vita to study the career paths of scientists and engineers: An exploratory assessment. *Scientometrics* 49(3): 419-442.

Dinopolous, E., and P. Thompson 1999. Reassessing the empirical validity of the human capital augmented neoclassical growth model. *Journal of Evolutionary Economics* 9: 135-154.

Edwards, B. M. and W. Foley 1998. Civil society and social capital beyond Putnam. *American Behavioral Scientist* 42: 124-139.

Elder, G. H. Jr. 1994. Time, human agency, and social change: perspectives on the life course. *Social Psychology Quarterly* 57: 4-15.

Elder, G. H. Jr., and E. K. Pavalko 1993. Work careers in men's later years transitions, trajectories, and historical change. *Journal of Gerontology* 48: S180-S191.

Faia, M. A. 1975. Productivity among scientists: A replication and elaboration. *American Sociological Review* 40: 825-829.

Faust, K. 1997. Centrality in affiliation networks. *Social Networks* 19: 157-191.

Friedkin, N. E. 1978. University social structure and social networks among scientists. *American Journal of Sociology* 83: 1444-1465.

Fusfeld, H. I. 1986. *The technical enterprise: present and future patterns*. Cambridge: Ballinger.

Gabbay, S. M., and E. W. Zuckerman 1998. Social capital and opportunity in corporate R&D: The contingent effect of contact density on mobility expectations. *Social Science Research* 27: 189-217.

Garson, G. D. 1998. Neural networks: An introductory guide for social scientists. London: Sage.

Goldberg, A. I., and Y.A. Shenhav 1984. R&D career paths: Their relation to work goals and productivity. *IEEE Transactions on Engineering Management* EM-31(3): 111-117.

Gomez-Mejia, L., and L. D. Balkin 1992. Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal* 35(5): 921-955.

Government Performance and Results Act of 1993. U. S. Public Law 103-62. 103rd Cong.

Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology* 78: 1360-1380.

Graves, S.B., and N.S. Langowitz 1996. R&D productivity: A global multi-industry comparison. *Technological Forecasting And Social Change* 53(2): 125-137.

Griliches, Z. ed., 1984. *R&D, patents, and productivity.* Chicago: The University of Chicago Press.

Griliches, Z. 1989. Patents—Recent tends and puzzles. *Brookings Papers On Economic Activity.* Sp. Issue: 291-330.

Griliches, Z. 1992. The search for R&D spillovers. *The Scandinavian Journal of Economics* 94: 29-47.

Gujarati, D. N. 1995. *Basic econometrics*. 3ed. New York: McGraw-Hill.

Gunz, H. P. 1980. Dual ladders in research: A paradoxical organizational fix. *R&D Management* 10: 113-118.

Gunz, H. P. 1989. The dual meaning of managerial careers: Organisational and individual levels of analysis. *Journal of Management Studies* 2: 225-250.

Heckman, J., and E. Vytlacil 1998. Instrumental variables methods for the correlated random coefficient model--Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources* 33: 974-987.

Heinsler, J., and R. Rosenfeld 1987. Charting academic careers. *Social Studies of Science* 17: 135-144.

Hummon, N. P., and K. M. Carley 1993. Social networks as normal science. *Social Networks* 15: 71-106.

Hunter, D. E., and G. D. Kuh 1987. The "Write Wing": Characteristics of prolific contributors to the higher education literature. *Journal of Higher Education* 58(4) 443-62.

Jaffe, A. B. 1989. Characterizing the 'technological position' of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy* 18(2): 87-97.

Jaffe, A. B., M. Trajtenberg, and R. Henderson 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108(3): 577-98.

Katz, J. S., and B. R. Martin 1997. What is research collaboration? *Research Policy* 26: 1-18.

Kayal, A. 1999. Measuring the pace of technological progress: Implications for technological forecasting. *Technological Forecasting and Social Change* 60(3): 237-245.

Keith, B., and N. Babchuk 1998. The quest for institutional recognition: A longitudinal analysis of scholarly productivity and academic prestige among sociology departments. *Social Forces* 76(4): 1495-1533.

Kennedy, P. 1998. *A guide to econometrics*. 4th ed. Cambridge, MA: MIT Press.

Kingsley, G., B. Bozeman, and K. Coker 1996. Technology transfer and absorption: An R&D value mapping approach to evaluation. *Research Policy* 25: 967-995.

Klette, T. J., and Z. Griliches 2000. Empirical patterns of firm growth and R&D investment: A quality ladder model interpretation. *Economic Journal* 110(463): 363-387.

Kostoff, R. N. 1994. Assessing Research Impact—Semiquantitative Methods. *Evaluation Review* 18: 11-19.

Kostoff, R., H. Averch, and D. Chubin 1994. Research impact assessment: Introduction and overview. *Evaluation Review* 18: 3-10.

Landry, R., N. Traore, and B. Godin 1996. An econometric analysis of the effect of collaboration on academic research productivity. *Higher Education* 32(3): 283-301.

Levin, S. G., and P. E. Stephan 1989. Age and research productivity of academic scientists. *Research in Higher Education* 30: 531-549.

Levin, S. G., and P. E. Stephan 1991. Research productivity over the life-cycle - evidence for academic scientists. *American Economic Review* 81: 114-132.

Long, J. S. 1978. Productivity and academic position in the scientific career. *American Sociological Review* 43: 889-908.

Long, J. S., P. D. Allison, and R. McGinnis 1979. Entrance into the academic career. *American Sociological Review* 44: 816-830.

Long, J. S., and R. McGinnis 1985. The effects of the mentor on the academic career. *Scientometrics* 7: 255-280.

Long, J.S., P. Allison, and R. McGinnis 1993. Rank advancement in academic careers: sex differences and the effects of productivity. *American Sociological Review* 58: 703-722.

Lotka, A. J. 1926. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16: 317.

Liyanage, S. 1995. Breeding innovation clusters through collaborative research networks. *Technovation* 15(9): 553-567.

MacRoberts, M. H., and B. R. MacRoberts 1986. Quantitative measures of communication in science: A study of the formal level. *Social Studies of Science* 16: 151-172.

McCormick, K. 1995. Career paths, technological obsolescence and skill formation—Research and development staff in Britain and Japan. *R&D Management* 25(2): 197-211.

Merton, R. K. 1957. Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review* 22: 635-659.

Merton, R. K. 1961. *Social theory and social structure*. Glencoe, IL: Free Press.

Merton, R. K. 1968. The Matthew effect in science. *Science* 159: 56-63.

Merton, R. K. 1993 [1965]. *On the shoulders of giants*: *The post-Italianate edition*. Chicago: University of Chicago Press.

Mincer, J. 1958. Investment in human capital and personal income distribution. *Journal of Political Economy* Aug: 281-302.

Mincer, J. 1997. The production of human capital and the life cycle of earnings: Variations on a theme. *Journal of Labor Economics* 15: S26-S47.

Mowery, D. C. 1998. The changing structure of the US national innovation system: Implications for international conflict and cooperation in R&D policy. *Research Policy* 27: 639-654.

Mullins, N. C. 1968. The distribution of social and cultural properties in informal communication networks among biological scientists. *American Sociological Review* 33: 786-797.

Nahapiet, J., and S. Ghoshal 1998. Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review* 23: 242-266.

Narin, F. 1976. Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity. *Contract report to the National Science Foundation*. Cherry Hill, NJ: Computer Horizons Inc.

Narin, F. 1994. Patent bibliometrics. *Scientometrics* 30(1): 147-155.

Narin, F. 1995. Patents as indicators for the evaluation of industrial research output. *Scientometrics* 34(3): 489-496.

Narin, F., and A. Breitzman 1995. Inventive productivity. *Research Policy* 24(4): 507-519.

Narin, F., and K.S. Hamilton 1996. Bibliometric performance measures. *Scientometrics* 36(3): 293-310.

Narin, F., K.S. Hamilton, and D. Olivastro 1997. The increasing linkage between US technology and public science. *Research Policy* 26(3): 317-330.

Narin, F., and D. Olivastro 1998. Linkage between patents and papers: An interim EPO/US comparison. *Scientometrics* 41(1-2): 51-59.

Narin, F., D. Olivastro, and K.A. Stevens 1994. Bibliometrics: Theory, practice and problems. *Evaluation Review* 18(1): 65-76.

(NAS) National Academy of Sciences 1999. *Improving student learning: A strategic plan for education research and its utilization*. Washington, D.C.: National Academy Press.

(NSB) National Science Board 2002. *Science and Engineering Indicators, 2002*. Arlington, VA: National Science Board.

(NSF) National Science Foundation 1978. The meaning of patent statistics. *Papers prepared for the National Science Foundation*. Washington, DC: National Science Foundation.

(NSF) National Science Foundation 1999. How much does the U.S. rely on immigrant engineers? *Issue Brief*. Arlington, VA: National Science Foundation.

(NSF) National Science Foundation 2001. *Characteristics of doctoral scientists and engineers in the United States: 2001 Profiles*. Arlington, VA: National Science Foundation.

Nelson, R. R. ed., 1993. *National innovation systems: A comparative analysis*. New York: Oxford University Press.

Nelson, R. R., and S. G. Winter 1982. *An evolutionary theory of economic change*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Nordhaug, O. 1993. *Human capital in organizations: Competence, training, and learning*. New York: Oxford University Press for Scandinavian University Press.

Patel, P., and M. Vega 1999. Patterns of internationalisation of corporate technology: Location vs. home country advantages. *Research Policy* 28(2-3): 145-155.

Pavitt, K. 1983. *Patterns of technical change: evidence, theory and policy implications*. London: Imperial College, University of London.

Pazy, A 1990. The threat of professional obsolescence: How do professionals at different career stages experience it and cope with it. *Human Resource Management* 29(3): 251-269.

Persson, O., and M. Beckmann 1995. Locating the network of interacting authors in scientific specialties. *Scientometrics* 33: 351-366.

Polanyi, M. 1967. *The tacit dimension*. London: Cox & Wyman.

Polanyi, M. 1969. *The logic of tacit inference—Knowing and being*. London: Routledge & Kegan Paul.

Price, D. J. d. S. 1963. *Little science, big science*. New York: Columbia University Press.

Price, D.J. d.S., and D. Beaver 1966. Collaboration in an invisible college. *American Psychologist* 21: 1011-18.

Putnam, R. D. 1993. *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.

Radosevic, S. and L. Auriol 1999. Patterns of restructuring in research, development and innovation activities in central and eastern European countries: an analysis based on S&T indicators. *Research Policy* 28(4): 351-376.

Rappa, M. A., and K. Debackere 1992.Technological communities and the diffusion of knowledge. *R&D Management* 22(3): 209-220.

Reskin, B. F. 1977. Scientific productivity and the reward structure of science. *American Sociological Review* 42: 491-504.

Reskin, B. F. 1978. Scientific Productivity, Sex, and Location in the Institution of Science. *American Journal of Sociology* 83: 1235-1243.

Roe, A. 1956. *The psychology of occupations*. New York: Wiley.

Roe, A. 1973. *The making of a scientist*. Westport, CT: Greenwood Press.

Rogers, E. M. 1995. *Diffusion of Innovations*. 4th ed. New York: The Free Press.

Rogers, J. D., and B. Bozeman 2001. Knowledge Value Alliances: An Alternative to the R&D Project Focus in Evaluation. *Science, Technology and Human Values* 26(1): 23-55.

Rosenberg, N., and R. R. Nelson 1994. American universities and technical advance in industry. *Research Policy* 23: 323-348.

Sarewitz, D. 1996. *Frontiers of illusion: Science, technology, and the politics of progress*. Philadelphia: Temple University Press.

Schmookler, J. 1966. *Invention and economic growth*. Cambridge: Harvard University Press.

Schultz, T. W. 1963. *The economic value of education*. New York: Columbia University Press.

Schultz, T. W. 1971. Investment in human capital. In B.F. Kiker ed*., Investment in human capital*. Columbia: University of South Carolina Press.

Scott, J. 2000. *Social network analysis: A handbook*. 2nd ed. London: Sage.

(SESTAT) Scientist and Engineers Statistical Data System 1999. *An online comprehensive and integrated system of information about the employment, educational, and demographic characteristics of scientists and engineers in the United States*. Arlington, VA: National Science Foundation.

Shepard, H. A. 1958. The dual hierarchy in research. *Research Management* 1: 177-187.

Simonton, D. K. 1990. Creativity in the later years—optimistic prospects for achievement. *Gerontologist* 30: 626-631.

Simonton, D. K. 1992. The social-context of career success and course for 2,026 scientists and inventors. *Personality And Social Psychology Bulletin* 18: 452-463.

Simonton, D. K. 1997. Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review* 104(1): 66-89.

Spagnolo, G. 1999. Social relations and cooperation in organizations. *Journal of Economic Behavior & Organization* 38: 1-25.

Stephan, P. E. 1996. The economics of science. *Journal of Economic Literature* 34: 1199-1235.

Stephan, P. E., and S. G. Levin 1992. *Striking the mother lode in science: The importance of age, place and time*. New York: Oxford University Press.

Stephan, P. E., and S. G. Levin 1997. The critical importance of careers in collaborative scientific research. *Revue d'Economie Industrielle* 79: 45-61.

Sudman, S., and N. N. Bradburn 1982. *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Sweetland, S. R. 1996. Human capital theory: Foundations of a field of inquiry. *Review of Educational Research* 66: 341-359.

Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26: 24-36.

Tsai, W. P., and S. Ghoshal 1998. Social capital and value creation: The role of intrafirm networks. *Academy of Management Journal* 41: 464-476.

Turpin, T., and A. Deville 1995. Occupational roles and expectations of research scientists and research managers in scientific research institutions. *R&D Management* 25(2): 141-157.

Walker, G., B. Kogut, and W. Shan 1997. Social capital, structural holes and the formation of an industry network. *Organization Science* 8: 109-125.

Wasserman, S., and K. Faust 1994. *Social network analysis*. Cambridge: Cambridge University Press.

Zucker, L. G., M. R., Darby, and J. Armstrong 1998. Geographically localized knowledge: Spillovers or markets? *Economic Inquiry* 36(1): 65-86.

Zuckerman, H. 1977. *Scientific Elite Nobel Laureates in the United States*. New York: Free Press.

VITA

Mr. Dietz was born in New York in 1966. He holds a Master of Public Administration degree (1989) in science policy from Syracuse University's Maxwell School for Citizenship and Public Affairs. He earned is baccalaureate (1988) in Public Policy and US History from the Rockefeller College of the University of Albany, State University of New York. In 1998 he enrolled in the doctoral program in Science and Technology Policy at the Georgia Institute of Technology where he is conducting dissertation research in the area of academic productivity and social and human capital theory.

Mr. Dietz currently works as an Associate Program Director in the Division of Research, Evaluation, and Communication of the National Science Foundation (NSF). His work spans all major programmatic areas of the Division. Mr. Dietz joined NSF in 1989 as a Presidential Management Intern—a federal government-wide program designed to train early career public administrators. During this two-year period, he worked in the NSF budget office, the program evaluation section, the Division of Mathematical Sciences, the Division of Research Initiation and Improvement, and the Office of the Assistant Director for Education and Human Resources. He also spent approximately eight months as a staff member in the US House of Representatives, Committee on Science and Technology where, among other duties, he authored legislation that was signed into law by President George H.W. Bush.