A New Taxonomy for Star Scientists: Three Essays

by

Alexander Oettl

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Management
University of Toronto

Abstract

A New Taxonomy for Star Scientists: Three Essays

Alexander Oettl

Doctor of Philosophy

Graduate Department of Management

University of Toronto

2009

It is surprising that the prevailing performance taxonomy for scientists (Star versus Non-Star) focuses only on individual output and ignores social behavior since scholars often characterize innovation as a communal process. To address this deficiency, I expand the traditional taxonomy that focuses solely on productivity and add a second, social dimension to the taxonomy of scientists: helpfulness to others. Using a combination of academic paper citations and Impact Factor-weighted publications to measure scientist productivity as well as the receipt of academic paper acknowledgements to measure helpfulness, I classify scientists into four distinct categories of human capital quality: All-Stars, who have both high productivity and helpfulness; Lone Wolves, who have high productivity but average helpfulness; Mavens, who have average productivity but high helpfulness; and Non-Stars, who have both average productivity and helpfulness.

The first study examines the impact of 415 immunologists on the performance of their coauthors. Looking at the change in quality-adjusted publishing output of an immunologist's coauthors after the immunologist's death, I find that the productivity of an All-Star's coauthors decreases on average by 35%, a Maven's coauthors by 30% on average, and a Lone Wolf's coauthors by 19%, all relative to the decrease in productivity of a Non-Star's coauthors. These findings suggest that our current conceptualization of star scientists, which solely focuses on individual productivity, is both incomplete and potentially misleading as Lone Wolves may be systematically overvalued and Mavens undervalued.

The second study builds upon the first study's finding that Mavens have a large impact on the performance of their coauthors. Using salary disclosures from 2008 at the University of California, I examine the extent to which each star type is compensated differently. While Mavens provide larger spillovers than Lone Wolves, Mavens are compensated less than Lone Wolves, providing preliminary evidence that the externalities generated by Mavens are unpriced.

The third study examines the likelihood of an immunologist's mobility as a function of his observable and unobservable human capital. The greater a scientist's productivity (observable to the market), the greater his inter-institution mobility, while the greater a scientist's helpfulness (unobservable to the market), the lower his inter-institution mobility.

Dedication

I dedicate this thesis to my parents, who taught me from an early age the value of hard work, the joy in learning, and to find beauty in the small things in life. For this, I am eternally grateful and so proud to be their son.

I also would like to dedicate this thesis to the love of my life, Karen. Her constant support, interest, willingness to “conference” my manuscripts, and above all love were invaluable in helping my through the pressures of doctoral studies. Without her, none of this would have been possible.

Acknowledgements

I would not have been able to finish this dissertation without the help of my dissertation committee chair, advisor, mentor, and friend, Ajay Agrawal. His sage advice and guidance as I meandered my way through graduate school and this dissertation are immeasurable. He has shaped me as a scholar, and I thank him for the often thankless (other than in dissertation acknowledgement sections) job of mentorship. I also would like to thank my dissertation committee members, Joanne Oxley and Olav Sorenson. Their support, help, and feedback have greatly improved both my scholarship and helped me grow as an academic.

The Rotman strategy group has provided me with an intellectually curious and above all friendly academic environment over these past five years. Every member of the strategy group has been helpful and encouraging; no one in the department is a Lone Wolf!

I additionally would like to thank my fellow PhD students for keeping me going when the going was tough and for helping to make my PhD experience a pleasant one: Bill Foster, Jay Horwitz, Nan Jia, Alison Kemper, Elena Kulchina, Alastair Lawrence, Sue Moon, and Paul Seaborn. In particular, I'd like to thank Christian Catalini, Colleen Stuart, and Diederik van Liere for not only (many) afternoon coffees but for being great friends and for helping me shape the questions I have attempted to answer in this study.

Lastly, this research was graciously funded by the Social Sciences and Humanities Research Council of Canada (Grant No. 410-2004-1770) and the Martin Prosperity Institute Program on Innovation and Creative Industries. Their support is greatly appreciated.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The innovation process is one of the most important aspects governing firm strategy. Increasingly, organizations of all stripes are placing a renewed emphasis on hiring "talent" to help them achieve their innovation goals (Guthridge, Komm, and Lawson, 2008). But what is talent? Largely, the term applies to highly skilled individuals, or stars. This focus stems from the empirical regularity that a small group of individuals (from almost any population) will generate a disproportionately large level of output. Bill Gates observed that "if it weren't for 20 key people, Microsoft wouldn't be the company it is today" (Gandossy, Tucker, and Verma, 2006, pp. 64)

It is surprising then that the prevailing dichotomous performance taxonomy for scientists (Star versus Non-Star) focuses only on individual output and ignores social behavior since, despite popular examples of Edison-like lone inventors, innovation is most often characterized as a communal process. The community benefits are largely due to common gains from specialization (innovation as the recombination of ideas) and the social mediation of knowledge flows (geographic proximity, ethnicity, and social networks mediate flows). Furthermore, knowledge transfer is famously difficult to contract for, heightening the value of conditions that favor the sharing of knowledge. Moreover, most strategy and economics research is less interested in innovative output at the individual level and instead focuses on aggregate output at the firm or regional level. This dissertation builds upon the current dichotomous conceptualization of star scientists by adding a second, social dimension to the taxonomy: helpfulness to others.

By adding this second dimension, the new taxonomy for star scientists is able to classify four distinct individuals: All-Stars are individuals with high personal productivity and high helpfulness, Lone Wolves are individuals with high personal productivity and average helpfulness, Mavens are individuals with average personal productivity and high helpfulness, and Non-Stars are individuals with average personal productivity and average helpfulness.

This dissertation thus poses three main questions. First, do the different star types in the new taxonomy for star scientists generate externalities (priced or unpriced) at differing rates for their coauthors? Second, if variation in generating externalities does exist across different types of star scientists, do the stars appropriate through salary the benefits they generate to their coauthors (the externalities are priced) or not (the externalities are unpriced)? Lastly, how does productivity and helpfulness influence the rate of mobility of star scientists to new institutions?

All three studies in this dissertation analyze multiple samples of academic immunologists (I examine academic economists in Chapter 3 as well). I focus on academic immunology for three reasons. First, it is a discipline characterized by similar environments to that of many bench scientists working at private-sector firms in such industries as pharmaceutical research, biotechnology, and semiconductors. Second, it is a discipline where frequent collaboration exists, critical for answering my first question, which examines the impact of different star types on the performance of their coauthors. Lastly, it allows for the measurement of the two main components of this new taxonomy for star scientists: personal productivity and helpfulness. I measure productivity using data on impact- and citation-weighted publication counts of scientists, very much in line with prior work. However, I measure helpfulness using previously unexploited data: acknowledgements from academic journal articles.

My first study, entitled *Productivity and Helpfulness: A New Taxonomy for Star Scientists*, examines the impact of a sample of 415 academic immunologists on the performance of their coauthors. I classify an immunologist as having high productivity and/or high helpfulness if he is in the top 5% of the sample's productivity or helpfulness distributions, respectively. To overcome the issue of endogenous coauthor matching (where immunologists self-select into coauthoring relationships, possibly on expected future performance), which obfuscates the estimation of a causal relationship between the presence of a tie between a star and coauthor, I examine the change in performance of coauthors when a star dies. By focusing on this semi-exogenous *cessation* in the coauthoring tie, I find that the productivity of an All-Star's coauthors decreases on average by 35%, a Maven's coauthors on average by 30%, and a Lone Wolf's coauthors on average by 19%, all relative to the decrease in productivity of a Non-Star's coauthors. These findings suggest that the current dichotomous conceptualization of star scientists is somewhat narrow as it classifies both All-Stars and Lone Wolves as stars, yet each has very different effects on the performance of their coauthors. Conversely, Mavens are currently classified as non-stars yet, as can be seen by their large impact on the performance of their coauthors, are presently unwisely overlooked as helpfulness stars. What cannot be determined from this study, however, is the extent to which the performance benefits conveyed onto the coauthors of stars are uncompensated and as such constitute unpriced externalities or spillovers.

The second study, *The Price of All-Stars, Lone Wolves, and Mavens* attempts to determine if the externalities generated by All-Stars, Mavens, and Lone Wolves are compensated (priced) or not (unpriced). Drawing on data of academic immunologists and economists in the University of California system, I generate a sample of 62 immunologists and 213 economists who received a salary in 2007. Because the University of California system is a public university system, all academic salaries are publicly available. Regressing a scientist's annual salary on the scientist's star type reveals that while Mavens provide larger spillovers (Chapter 2) than Lone Wolves, Mavens are compensated less than Lone Wolves, providing preliminary evidence that the externalities generated by Mavens appear to be unpriced and thus constitute spillovers. Furthermore, Mavens who engage in more internal collaboration, thus allowing the *Maven*'s institution to capture more of the spillovers generated by him, are paid more than Mavens who collaborate more with coauthors outside of the institution.

The third study, *Asymmetric Information and Labor Mobility: The Case of Scientist Productivity*, examines the extent to which the underlying characteristics of productivity and helpfulness either facilitate or impede scientist mobility between institutions. Building on the results presented in the first two studies, both productivity and helpfulness provide performance benefits to institutions. However, while a scientist's productivity is observable to outside institutions, which increases the scientist's value and thus has a positive impact on the scientist's mobility, a scientist's helpfulness is only observable to the scientist's own institution. As a result of this asymmetry in observing the true human capital of a scientist for outside institutions, I theorize that the value of helpful scientists is more uncertain, which should have a negative effect on their mobility. The findings on a sample of 4,665 academic immunologists between 1974 and 1989 show that indeed the more productive a scientist is the more likely he is to move, yet the more helpful a scientist is the less likely he is to move. This study shows that the underlying characteristics of stars who generate the largest spillovers (in other words, helpfulness) is also a source of immobility, not out of choice, but because of a market failure in pricing the characteristics that would induce a scientist to move. As a result, helpfulness stars (All-Stars and Mavens) can be sources of competitive advantage due to their imperfect labor mobility.

These three studies attempt to expand our thinking of what a star scientist is. High productivity scientists do not generate the largest externalities for their coauthors, while high helpfulness scientists do produce the largest externalities for their peers. As such, prior work, which has solely focused on the personal productivity dimension of an individual, has been both incomplete and misleading in characterizing the performance benefits of stars.

Peteraf (1993, pg. 187) wrote that "a Nobel Prize-winning scientist may be a unique resource, but unless he has firm-specific ties, his perfect mobility makes him an unlikely source of sustainable advantage." Peteraf is correct. High productivity stars (Lone Wolves) may not be sources of sustainable advantage, yet this dissertation demonstrates, that a separate type of scientist, high helpfulness stars, (Mavens) *can* act as a source of sustainable advantage for three reasons. First, they generate performance benefits to their peers in addition to their own personal productivity. Second, the market does not fully price these performance benefits so that they are indeed spillovers. Third, the more helpful an individual is, the less mobile he is, allowing organizations that do employ Mavens to utilize them as a source of sustainable advantage.

Chapter 2

Productivity and Helpfulness: A New Taxonomy for Star Scientists

2.1 Introduction

The need to hire the best and the brightest - "the war for talent" - has long been one of the most pressing strategic concerns facing managers (Kapur and McHale, 2005; Guthridge, Komm, and Lawson, 2008). This concern is largely driven by the observation that high performers, or stars, account for the generation of a disproportionately large level of output. Google's vice-president of engineering, Alan Eustace, noted to the Wall Street Journal in 2005 that "one top-notch engineer is worth 300 times or more than the average" and that he "would rather lose an entire incoming class of engineering graduates than one exceptional technologist" (Tam and Delaney, 2005). Why is this? How do stars so greatly influence the performance of organizations?

The existing performance taxonomy for scientists focuses exclusively on individual output, classifying a scientist as either a Star or a Non-Star. The seminal work of Zucker, Darby, and Brewer (1998), for example, defines stars as the top 0.75% of contributors to the genetic sequence database GenBank, a group that accounts for almost 17% of contributions. Recent work by Groysberg, Lee, and Nanda (2008) examines the skill portability of the top 3% of security analysts when they move firms using a ranking of the perceived effectiveness of security analysts, while Azoulay, Graff Zivin, and Wang (2008) look at the impact of eminent scientists using a variety of measures such as research funding, citations, and patenting. In all of these articles, the definition of a star is based solely on individual productivity; in other words, we define stars by what they physically produce.[1]

This uni-dimensional classification of star scientists is surprising as innovation is most often characterized as a communal process. Communal interactions matter for two reasons. First, innovation is more often a result of the recombination of existing knowledge and ideas rather than the discovery of something fundamentally novel (Gilfillan, 1935; Nelson and Winter, 1982). As knowledge frontiers continue to expand, combinations of increasingly specialized levels of human capital are required to reach the forefront of knowledge (Wuchty, Jones, and Uzzi, 2007; Jones, 2009). It is this recombination of specialized ideas, either through formal collaborations (coauthorships, joint ventures, etc.) or informal means (discussions and comments from helpful individuals), that leads to innovation. Second, the exchange of knowledge is to a large extent governed through social channels. Individuals possess only finite levels of knowledge, and knowledge search is costly. Social forces can reduce barriers to knowledge flow through geographic proximity (Jaffe, Trajtenberg, and Henderson, 1993), labor mobility (Almeida and Kogut, 1999; Oettl and Agrawal, 2008; Singh, 2005), and membership in ethnic communities (Agrawal, Kapur, and McHale, 2008).

While innovation is a communal process, the inability for parties to perfectly contract on knowledge exchange leads to failures in the market for knowledge and a decrease in knowledge transfer (Arrow, 1962). As such, conditions that facilitate knowledge sharing

---

[1] While I am unable to directly measure individual productivity, I assume that an individual's inputs are uniform and constant and thus use the term productivity as a measure of an individual's output.

or spillovers in the absence of formal contractual environs are of great value to firms. Ultimately, if our concern is to understand the mechanisms by which an individual maximizes his performance, simply understanding the productivity inputs of an individual would suffice. However, the strategy and economics literatures focus on performance measures at the organization and regional levels, and as such, mechanisms in which individuals influence the productivity of others become important as these mechanisms directly influence the performance of organizations and regions. Hence, mechanisms by which individuals generate spillovers are of paramount concern to scholars of strategy and economics.

The importance of social factors on innovation illuminates the deficiency of our current productivity-focused conceptualization of star scientists (Stars versus Non-Stars). To expand our current conceptualization of star scientists, I develop a new taxonomy of star scientists by incorporating a social dimension: helpfulness to others. This new taxonomy allows an individual to not only vary along a productivity dimension but also a helpfulness dimension.

The objective of this chapter is threefold. First, I expand upon the current dichotomous conceptualization of stars by developing a taxonomy that not only incorporates a star's individual productivity but also his helpfulness. In doing so, I move beyond the current uni-dimensional classification and redefine what it means to be a star. Second, I propose a measure to classify individuals into this new taxonomy. Third, I use this taxonomy to assess the extent to which different star types influence the productivity of others.

Following prior studies (Allison and Long, 1990; Azoulay, Graff Zivin, and Wang, 2008), I measure individual productivity using Impact Factor-weighted publication counts.[2]

---

[2] The Impact Factor is a time-varying journal-level measure of quality, which captures the rate at which each article in the journal is cited. Thus, journals with articles that are cited more often will have higher Impact Factors. The Institute for Scientific Information (ISI), a subsidiary of Thomson Scientific, constructs this measure annually.

On the other hand, I measure helpfulness by academic journal acknowledgements, since such acknowledgements are generally made to those who have helped in the development of the work. Using these measures of productivity and helpfulness, I classify a sample of 415 immunologists and examine their influence on the productivity of their coauthors. I use coauthorship to pinpoint the timing of the formation of an interpersonal tie between an immunologist and a potential recipient of spillovers. It is this co-location in social space that allows the coauthor the potential to benefit from any spillovers the star may provide.[3]

By placing a star in both productivity and helpfulness space while keeping the classifications discrete, I am able to classify an individual as one of four types: All-Star, Lone Wolf, Maven, or Non-Star.

I define an All-Star as an individual with both high productivity and high helpfulness. A Lone Wolf is someone who has high productivity but average helpfulness. A Maven is an individual with average productivity but high helpfulness, and a Non-Star has both average productivity and average helpfulness. Restrictively, the current dichotomous conceptualization of stars groups both All-Stars and Lone Wolves together, while completely overlooking Mavens. By expanding on the current classification, I am able to examine the influence of individuals who vary in both their productivity and helpfulness.

Examining the changes in productivity from coauthoring with various star types would be an appropriate empirical exercise if coauthoring relationships were chosen at random, but clearly they are not. The problem with endogenous coauthor selection is that the coauthors selected by an immunologist may be chosen due to their own productivity, thus producing spurious correlations between an individual's productivity and their coauthorship network. For this paper, I examine the *decrease* in productivity of coauthors when an immunologist dies.

---

[3] To be clear, this study cannot make the claim that these spillovers are externalities as pricing data are not available for the sample studied.

Across a number of specifications, the productivity of the coauthors of All-Stars who die decreases on average by 35% relative to the decrease in productivity when a Non-Star dies. More interestingly, coauthors of Mavens who die experience a 30% decrease in productivity, while the coauthors of Lone Wolves who die experience decreases in productivity of only 19% on average.

By expanding the current conceptualization of star scientists and focusing on both the productivity and helpfulness dimensions of scientists, I find that spillovers are most likely generated from individuals with high helpfulness. As a result, the literature has largely overemphasized the importance of Lone Wolves yet has overlooked and consequently underemphasized Mavens.

2.2 Star Scientists and Spillovers

While the quality of human capital may be uniformly distributed, the returns to human capital are wildly skewed (Ernst, Leptien, and Vitt, 2000). Individuals in the right tail of the distribution, so-called stars, generate a disproportionately large share of output (Rosen, 1981). As Lotka (1926) observed, the top 6% of physicists produce more than 50% of all papers. This skewed distribution - termed the Pareto Principle - is ubiquitous across industries and is a strong determinant of inventive productivity (Narin and Breitzman, 1995). However, as I argue, even though stars contribute disproportionately to output production, they cannot alone act as a source of sustained competitive advantage.

In the resource-based view (RBV) of the firm, firms generate sustainable competitive advantage through their use of strategic resources (Wernerfelt, 1984). Resources are only sources of sustainable competitive advantage, however, if they are valuable, rare, inimitable, and difficult to substitute (Barney, 1991). In efficient factor markets, a star will be perfectly compensated for his productivity (Hirshleifer, Glazer, and Hirshleifer, 1998), suggesting that they cannot be a source of rents for the firm. The resource-picking literature, nonetheless, theorizes that firms can capture economic rents by employing superior information or analysis to pick undervalued resources in the factor market, much in the same way that a fund manager attempts to outsmart the financial markets by picking stocks (Barney, 1986; Makadok, 2001). The central assumption of this resource-picking mechanism is that the factor markets must be characterized by imperfect information, thus providing a forum for a firm with superior information to pick undervalued resources. While this situation may well exist in a number of factor markets, it seems unlikely to exist in the market for stars, as the output of stars is both easy to measure and highly visible. Consequently, the likelihood of a star being mispriced is low, rendering the resource-picking mechanism ineffective at generating sustainable competitive advantage. Furthermore, because of the high visibility of the output of stars, stars themselves are more mobile, further attenuating the value of stars as resources (Lazear, 1986).

Human capital, however, can be important for firm strategy in ways other than generating direct output: human capital can generate spillovers. Since the early work of Lucas (1988) human capital spillovers have been at the center of economic growth models. Lucas classifies human capital into two types: internal and external. Internal effects of human capital capture the extent to which human capital affects the individual's own productivity, while the external effects of human capital capture the influence individuals have on the performance of others. If these external effects generate an unpriced spillover onto the productivity of others, then the spillover constitutes an externality (Acemoglu, 1996). The notion that these human capital externalities and their effect on the increase in knowledge stocks can lead to increasing economic returns is captured by the endogenous growth theory of Romer (1990). As knowledge flows and spillovers lie at the center of many of our models of innovation (Audretsch and Feldman, 1996) and as human capital externalities are a key input in the generation of knowledge flows, understanding the parameters within which human capital spillovers are generated is of utmost importance. In addition, from a firm strategy standpoint, human capital spillovers may be more difficult to observe, thus allowing for the possibility of resource picking and conversely being used human as a source of sustainable competitive advantage.

Despite the importance of human capital spillovers, the strategy and economics literature has mostly focused on the skewed nature of the productivity distribution when examining the relationship between stars and performance. The seminal work of Zucker, Darby, and Brewer (1998) reports strong correlations between the location of star scientists and the formation of biotechnology ventures. In more recent work, Groysberg, Lee, and Nanda (2008) examine the firm specificity of the human capital of stars. They find that much of the performance premiums accruing to stars is firm specific and that when stars move, their productivity decreases. In a related work, Brown (2008) explores the relationship between effort exertion and differences in relative ability. Using a sample of professional golfers, Brown finds that golfers exert less effort in the presence of a star (Tiger Woods), thus indicating that the presence of stars can have a negative effect on organizational outcomes. None of these studies, however, explicitly examines these stars' human capital spillovers.

One notable and important exception is the work of Azoulay, Graff Zivin, and Wang (2008), henceforth referred to as AGW. AGW examine the effect that the death of an eminent life scientist has on the performance of his coauthors.[4] They find that, following the death of a star, coauthors' productivity decreases by 5% to 10%, interpreting this decrease in productivity as a sign of the presence of spillovers. Since spillovers were occurring between the star and the coauthor, the cessation of the coauthoring relationship (because of the death) ended these spillovers, resulting in a decline in coauthor productivity. Their study, however, is unable to rule out two possibly conflating effects. First, because their sample consists only of the top 5% of life scientists who died and their analysis solely examines the influence of those deaths on the productivity of their coau-

---

[4] AGW classify a scientist as eminent if he matches a number of performance-related criteria. In general, one can view these life scientists as being in the top 5% of the life scientist productivity distribution.

thors, they are unable to examine the effect of spillovers from Non-Stars, as Non-Stars are not included in the sample. As such we are unable to determine how much greater the spillovers are from stars than Non-Stars. They do show evidence that spillovers are increasing in a scientist's citations, but this is still conditioned on being in the top 5%. Second and in a related vein, without the inclusion of different star types in their sample, they are unable to disentangle the decrease in productivity of coauthors that is due to the loss of an intellectual link versus the decrease in productivity arising from the emotional toll of the death of a former colleague.

This paper builds upon and extends the pioneering work of AGW by developing a new taxonomy of star scientists that allows for more precise identification of star scientists of different productivity types and examines the conditions under which stars are more likely to impact the productivity of their coauthors.

2.3 A New Taxonomy of Star Scientists

The difficulty in finding empirical examples of human capital externalities is due to the uneven distribution of the externality generating process. In others words, not all individuals produce equal levels of human capital spillovers. The current strategy and economics literature classify stars along a single dimension - productivity. That is, an individual is classified a star if he falls in the right tail of some productivity distribution, normally output. For example, Zucker, Darby, and Brewer (1998) classify the top 0.75% of GenBank contributors as stars. Yet if social behaviors influence the spillover-generating process, then including a dimension of social behavior in our conceptualization of star scientists is surely needed. I extend our current conceptualization of star scientists that solely focuses on productivity and add a second, social dimension to the taxonomy of scientists: helpfulness. Where productivity encapsulates an individual's output that is beneficial to himself, helpfulness encapsulates an individual's output that is beneficial to others.

Examining the helpfulness of individuals in organizations is a well-trodden research stream. A large amount of literature on organizational psychology examines what is known as Organizational Citizenship Behavior (OCB) (Smith, Organ, and Near, 1983). The literature finds that a combination of altruism and courtesy greatly influences the level of helpfulness individuals extend to one another within organizations. A large literature in social psychology exists on the personality characteristics associated with helpful behavior. Among the many factors that influence an individual's helpfulness, three are most applicable to the setting of academic scientists: situational, social, and person factors. Situational factors deal with the costs associated with helping, social factors involve the influence of social norms on helpful behavior, and person factors capture the prosocial traits of an individual. I believe that holding situational and social factors constant within my setting of academic scientists is acceptable, and as such, person factors (which are innate) should be the only traits that influence helpfulness within my study (Fletcher and Clark, 2003). Consequently, I take helpfulness to be an innate, continuous performance measure that captures an individual's output that is beneficial to others.

To provide a more concrete description of what helpfulness in the sciences entails, below are some quotes from the obituaries of Maurice Landy and D. Bernard Amos, two scientists who are members of my empirical sample and both of whom are in the top 5% of the helpfulness distribution.

On Maurice Landy:

> "During his last year, Maurice Landy was a member of my laboratory, where he gave generously of his wisdom and experience. He was particularly attentive to younger scientists, teaching them to present their work in its optimal light and to respect and critically enjoy the work of others. Friends are valued for much more than just their contributions to knowledge. Thusly, Maurice,

> as a friend, turned our failures into successes, our parochialisms into worldliness, and our desperations into hopes. He demanded the best out of us and enjoyed our accomplishments as he might his own. And, in the end, he left a memory that we all cherish." (Lawrence and Cohn, 1993)

On D. Bernard Amos:

> "Bernard has had a profound impact on many individuals during his life. He has been instrumental in the training, education and development of generations of clinical and basic scientists." (Tedder and Dawson, 2003)

and

> "Only two months before his death I asked him about a research problem I had encountered testing individuals exposed to tuberculosis that have a negative skin test. He remembered his NLT work of 40 years earlier and suggested that I look for antibodies to Class II and antibodies to tuberculin that could prevent delayed hypersensitivity reactions tested by intradermal inoculation of tuberculin." (Yunis, 2004)

In comparison, the obituaries of Zanvil Cohn and Philip Gell, two immunologists in the top 5% of the productivity distribution but not the top 5% of the helpfulness distribution, focus more on the scholastic accomplishments of the two scholars. On Zanvil Cohn:

> "His incisive manner, his admiration of clever new experiments, his sense of fairness and respect, and his wit all will be sorely missed." (Steinman and Moberg, 1994)

On Philip Gell:

> “He was one of the founders of the British Immunological Society and started the first postgraduate M.Sc. course in immunology in 1963, which is still running today. In this course, and with a stream of postdoctoral fellows in his own laboratory, Gell helped to influence many of the basic and clinical scientists, both British and international, who lead the field today. He was elected Fellow of the Royal Society in 1969. (Silverstein and Benacerrafe, 2001)

Table 2.1 presents a new taxonomy for star scientists that incorporates productivity and helpfulness. Not only does a scientist vary along a dimension of productivity, he also varies along a measure of helpfulness. In doing so, I define three new star types. An All-Star is an individual with both high productivity and helpfulness. A Lone Wolf is an individual with high productivity but average helpfulness. A Maven is an individual with average productivity but high helpfulness. A Non-Star, has both average productivity and helpfulness.

Why does this taxonomy matter? Conventionally, both All-Stars and Lone Wolves are classified as stars as they both have high productivity. This aggregation has large strategy implications if the effects of All-Stars and Lone Wolves on organizations vary. Mavens on the other hand are currently classified as individuals with average productivity. But Mavens may provide some of the largest spillovers to others due to their level of helpfulness. As such, we may be overvaluing Lone Wolves while undervaluing Mavens. Given that human capital spillovers are at the core of our innovation and economic growth models, it is paramount to identify which inputs into the economic production function generate the largest spillovers.

2.4 Data

An ideal empirical setting for this study satisfies three criteria. First, it should take place in an organizational setting where collaboration exists. As the goal of this paper is to identify which types of individuals generate the largest spillovers for their coauthors, a setting in which spillovers take place is clearly necessary. Second, from a measurement standpoint, the ability to separate individual from group or organizational level performance is necessary as the focus of interest is on star individuals and not star teams or firms. Third, a field or discipline that engages in the practice of manuscript acknowledgements is necessary to identify individual helpfulness, which I will discuss in more detail later in this dissertation. A discipline that satisfies all three of these conditions is the field of immunology.

From a research standpoint, immunology is an incredibly important discipline. The National Institute of Allergy and Infectious Diseases (NIAID), which oversees the distribution of immunology-related research grants, allocated \$940 million to immunology research in 2005, up from \$646 million in 2003 (Hackett, Rotrosen, Auchincloss, and Fauci, 2007). More importantly, however, the structure of immunology research is organized in a very similar fashion to other medical sciences, such as biochemistry, microbiology, and pharmacology.

2.4.1 Measures

One major hurdle to extending the dichotomous conceptualization of stars has been the lack of data. I propose to use the receipt of acknowledgements as a measure of an individual's helpfulness. Academic acknowledgements are a central and convenient way of recognizing a non-author's contributions to the development of a manuscript without extending ownership rights in the form of coauthorship (Merton, 1973).[5]

---

[5] Of course, acknowledgements can come in two forms. They may represent an acknowledgement of another author's useful comments (that is, the author is selected on quality) or they may accrue as

The goal of this study is to classify an immunologist along the dimensions of productivity and helpfulness and then to examine the change in output of his coauthors when he dies, thus inferring the presence of spillovers. I measure productivity as the total number of citations received for papers written by the focal immunologist prior to 1966. Citation data come from the Institute for Scientific Information (ISI) Web of Science. I measure helpfulness as the total number of acknowledgements received by the focal immunologist between the years 1960 and 1965 (inclusive) in *The Journal of Immunology.* I choose *The Journal of Immunology* because during this time period it was the pre-eminent academic journal for the discipline of immunology.[6] Acknowledgements operate very similarly in immunology as they do in the social sciences, albeit with fewer acknowledgements per paper. Of the 1,324 articles published in *The Journal of Immunology* between 1960 and 1965, 50% had at least one acknowledgement. Of the articles that did have at least one acknowledgement, 40% of them had at least one acknowledgement for criticism and encouragement, the measure used for this study.[7] As an example, the following was in the acknowledgement section of Bennett (1965): "The author wishes to thank Drs. L. J. Old and E. A. Boyse of the Sloan-Kettering Institute, New York, for their suggestions and encouragement, and Mrs. Patricia Hubertus for technical assistance."

I measure a coauthor's productivity by their Impact Factor-weighted publications. I obtain Impact Factor weights from the Journal Citation Reports from the ISI, which published Impact Factors for all immunology journals on a yearly basis between 2000 and

---

a result of the author's influence on the publishing process, the field, etc. (the author is selected on status). While I am unable to empirically separate out these two types of acknowledgements, "status" acknowledgements should add noise to the empirical analysis and thus, due to attenuation bias, result in conservative estimates. I discuss this further in Section 2.8.

[6] *The Journal of Immunology* in 2007 had an Impact Factor of 6.068, ranking it 13th among all immunology journals. It is, however, by far the most widely cited journal in immunology and has been in print since 1916, making it one of the oldest immunology journals in the world. Furthermore, the *Journal of Immunology* was chosen at random, and I have no reason to believe that immunologists providing feedback or criticism alter their behavior on publications intended for the *Journal of Immunology.*

[7] Acknowledgements thanking lab technicians and assistants are removed, although my sampling process requiring at least three lifetime publications most likely would have removed these individuals anyway.

2007. I use the average Impact Factor across these eight years to create a time-invariant quality measure of the 136 immunology journals indexed by ISI.

I collect data on deaths in a hybrid form by way of extracting obituaries and memoriams from the titles of over 400,000 immunology articles from the Web of Science as well as through manual Internet searches. While ideally I would like to identify unexpected deaths so that the "treatment" of losing a coauthor is fully exogenous, none of the deaths in my sample are of this nature and as such may not be fully exogenous. Allowing for the possibility that the deaths were anticipated by coauthors should generate conservative estimates of the productivity effect, as presumably the coauthors had time to make alternate arrangements to minimize the anticipated decrease in productivity. As such, the regression estimates should be viewed as the changes in productivity net of an anticipation of death.

2.4.2 Sample

The sample for this study draws on all immunologists who published at least one article in an immunology journal between 1960 and 1965, inclusive.[8] There are 5,323 of these scientists. I apply the requirement constraint that every immunologist must have at least three lifetime papers in an attempt to remove post-doctoral students, graduate students, and any other scientists who did not become academic scientists. This reduces my sample to 1,543 scientists. This remaining set of scientists must meet one final criterion: they must have at least one coauthoring relationship, formed after 1965, with a scientist who also has at least three lifetime publications. After applying this final condition, I am left with a final sample of 415 immunologists.

I divide my sample of 415 immunologists into four discrete categories based on their

---

[8] I draw the list of immunology journals from the Thomson Corporation's ISI Web of Science database. While I include the major immunology journals, such as *The Journal of Immunology* and *The Journal of Experimental Medicine*, in this list, general field journals such as *Nature* and *Science* are not. Consequently, I do not use "general" journals when constructing measures and defining the sample.

location in productivity and helpfulness space. Figure 2.1 graphically shows the sample's placement. While the constructs of productivity and helpfulness are continuous measures, the new taxonomy for star scientists requires discrete allocations and as such a cut-off point must be established to discern between high and average levels of productivity and helpfulness. The goal of this taxonomy is not to quibble about cut-off points but rather to demonstrate that an individual at the extreme end of one distribution is of a different type than the average individual in the rest of the distribution. Therefore, I define high productivity and helpfulness as being in the top 5%, which is a similar cut off to other studies looking at stars (Azoulay, Graff Zivin, and Wang, 2008). An immunologist has high productivity (in the top 5%) if he receives more than 2,028 citations prior to 1965, and an immunologist has high helpfulness (in the top 5%) if he receives three or more acknowledgements in *The Journal of Immunology* in the six years between 1960 and 1965, inclusive.

All-Stars are immunologists who have both high productivity and helpfulness (upper right quadrant of the graphic). Mavens are immunologists with average productivity but high helpfulness (upper left quadrant). Lone Wolves have high productivity but average helpfulness (bottom right quadrant). Non-Stars have both average productivity and helpfulness (bottom left quadrant). I show the classification of the sample in tabular format in Table 2.2.

Assigning the sample of 415 immunologists results in the following classification: four scientists are All-Stars, five are Mavens, 16 are Lone Wolves, and 390 are Non-Stars.[9] Of these 415 immunologists, 28 of them have died: two All-Stars, three Mavens, five Lone Wolves, and 18 Non-Stars.

---

[9] The observant reader will notice that while the high productivity stars (All-Stars and Lone Wolves) account for 5% of the total sample ($\frac{4+16}{415}$), the high helpfulness scientists do not ($\frac{4+5}{415}$). The reason for this is that because of the discrete nature of acknowledgements and the requirement that membership in the top 5% be *greater than* the 95th percentile cut-off values, I classify fewer than 5% of all scientists as highly helpful. When I change the cut-off designation to define scientists in the top 5% as scientists with *greater than or equal to* the 95th percentile cut-off values, the results are statistically and quantitatively largely unchanged.

2.4.3 Unit of Analysis

To what extent do different star types influence the productivity of others? To answer this question, I look at the change in productivity of coauthors of stars who die. As such, my unit of analysis is an immunologist-coauthor-year triad. The cross-sectional unit, however, is the immunologist-coauthor dyad, where the immunologist is one of four star types.[10] To identify coauthors, I identify all coauthorships formed after 1966 with scientists who have at least three lifetime publications. The immunologists in the sample have 58.3 coauthors on average, resulting in 24,175 immunologist-coauthor dyads.[11] The average publishing lifespan for immunology coauthors in my sample is 23.8 years, resulting in a final sample size of 575,483 observations. For the 28 immunologists who die, I reduce the sample to 816 dyads generating a subsample of 25,968 observations.

Using 1965 as a cutoff for both the productivity and helpfulness measures allows me to hold constant each scientist's "type" for 1965.

While the new taxonomy for star scientists is meant to classify an individual at a given point in time, for this paper I assume that distributions of skill along both the productivity and helpfulness dimensions are innate and thus do not vary across my sample. By looking at all coauthorships formed after 1965, I hold the window of evaluation constant for all star types and only focus on a star's influence on the productivity of all new coauthors, thus reducing the likelihood that an immunologist's productivity and helpfulness is a function of the coauthor's productivity. In addition, by only looking at newly formed coauthorships after immunologist types have been established in 1965, I remove the conflation of star definition with the performance of coauthors in that I only

---

[10] Non-Stars, the fourth type of star, are of course not truly stars, but for ease of classification I will consider them as one of the four star types.

[11] What happens if a scientist is the coauthor of multiple focal immunologists? While 78% of the coauthors in the sample only coauthor with one of the focal 415 immunologists, 15% coauthor with two immunologists in the sample, 4% coauthor with three immunologists in the sample, and 3% coauthor with four or more immunologists in the sample. I can adjust standard errors to account for this serial correlation. In practice, however, standard errors that have been adjusted for both immunologist and coauthor serial correlation differ marginally from standard errors adjusted solely for immunologists.

examine the performance impact of scientists who are not coauthors of the star at the time of classification. Furthermore, to reduce the conflation of star type classification with the productivity lifecycle of scientists, I include various age group dummies for both immunologists and coauthors in all regression models. I further discuss estimation and control variables in Section 2.5.

2.4.4 Descriptive Statistics

Tables 2.3 and 2.4 present descriptive statistics and correlation matrices for both the full sample of 24,175 dyads and the subsample of 816 dyads where the focal immunologist dies. For the full sample (Table 2.3), the average coauthor publishes 9.737 Impact Factor weighted publications a year, while the average coauthor of an immunologist who dies publishes on average 11.029 Impact Factor weighted publications a year. Table 2.5 presents means of four performance measures split both by the type of star the focal immunologist in the dyad is and whether or not the immunologist died. The coauthors of All-Stars and Lone Wolves have higher average Impact Factor weighted papers than Mavens and Non-Stars, but the coauthors of Mavens receive, on average, the most citations for their papers. Across almost all star types and measures, the coauthors of immunologists who die have higher output, on average, than the coauthors of immunologists who do not die.

2.5 Econometric Estimation

The empirical objective of this study is to examine the extent to which different star types influence the productivity of others. As discussed earlier, a star has the ability to influence the productivity of individuals across multiple levels: coauthors, peers in the same department, peers within the same institution, etc. For this study, I solely focus on a star's influence on the productivity of his coauthors.[12] The most straightforward empirical approach would be to examine the change in a coauthor's productivity after the formation of the coauthoring relationship (i.e., after the first time the two scientists collectively author a paper). Unfortunately, both the decision to coauthor at all and the decision of whom to coauthor with are clearly not random decisions. This endogeneity would bias my regression coefficients as the choice of coauthors may very well be related to their future productivity, resulting in a spurious relationship between a coauthor's productivity and his coauthorship network. As such, the empirical challenge becomes finding an exogenous change in the coauthoring relationship. An alternative to examining the formation of coauthoring ties is to examine the cessation of coauthoring ties but one that is exogenous. For this paper, I examine the change in productivity of a coauthor when an immunologist dies.

The empirical model to be estimated is

$$\begin{aligned} Y_{-ijt} = exp[&\beta_1 Death_{it} + \beta_2 Death_{it} \times AllStar_i + \beta_3 Death_{it} \times LoneWolf_i + \\ &\beta_4 Death_{it} \times Maven_i + \gamma_{it} + \mu_{jt} + \delta_t + \phi_{ij} + \varepsilon_{ijt}] \quad (2.1) \end{aligned}$$

Since my objective is to capture the change in productivity of a coauthor after an immunologist's death, the dependent variable, $Y_{-i,j,t}$, measures the number of Impact Factor weighted publications coauthor j wrote in year t where star i is *not* a coauthor. I use quality adjusted publication counts instead of raw publication counts to ensure that I am observing changes in the quality of publishing rather than changes in the frequency of publishing. $Death_{ijt}$ is an indicator variable that switches to 1 the year immunologist i dies. β_1 captures the net change in productivity of coauthor j after star i dies, irrespective of his star type. β_2, β_3, and β_4 captures the change in productivity of coauthor j if

---

[12] In a working paper, Waldinger (2008) explores changes in the productivity of scientists (peer effects) from the exogenous dismissal of colleagues in Nazi Germany. He explores peer effects at three levels: the department level, the same specialization within the same department, and the coauthor level. He only finds evidence of peer effects at the coauthor level.

immunologist i is an All-Star, Lone Wolf, or Maven, respectively. I omit the Non-Star category, and so the coefficients of β_2, β_3, and β_4 should be interpreted as the change in productivity relative to the productivity change when a Non-Star dies. Because the star types of i are time invariant, they can only be identified through the interaction with *Death*. γ_{it}, and μ_{jt} are sets of age cohort dummies that capture the changes in research productivity across the academic lifecycle (Levin and Stephan, 1991).[13] I capture time effects with δ_t. ϕ_{ij} is a series of dyad fixed effects, which in practice are conditioned out during estimation and as such are not directly estimated. ε_{ijt} is an identically distributed error term but not independent. Errors are correlated due to star i's death affecting all of his coauthors at the same time. Clustering of the standard errors by the star will correct for this non-independence during estimation. If the coefficients on β_1 through β_4 are less than zero, then the death of star i has a negative influence on the productivity of coauthor j, which provides some evidence that star i is influencing the productivity of coauthor j.

The identification of $Death_{ijt}$ comes from the variation in the deaths of immunologist i. By employing dyad fixed effects, I capture all time invariant attributes common to the dyad by these fixed effects, forcing the parameters to be solely identified from within dyad variation. Because of the count nature of the dependent variable and the high percentage of zero values (33%) across the sample, a count model is most appropriate. Specifically, I employ the Fixed Effects Poisson (FEP) estimator developed by Hausman, Hall, and Griliches (1984). Apart from being computationally straightforward, the Fixed Effects Poisson estimator estimated via quasi maximum likelihood (QML) has strong robustness features, even allowing for consistent parameter estimates of non-count dependent variables (Wooldridge, 2002). In addition, standard errors can be made robust to deviations from the poisson distribution, in particular the equality requirement of the

---

[13] In practice, I generate these age cohort dummies in four-year intervals, whereby the first dummy will capture a scientist in his first four years, the second dummy will capture a scientist in his fifth through eighth year, etc.

first and second moments (Wooldridge, 1999). I report these robust standard errors for all QML specifications.

In addition to the QML Fixed Effects Poisson specification used, I implement a linear model of the following form for robustness checks:

$$\begin{aligned} log\tilde{Y}_{-ijt} = d_{ijt} + \beta_1 Death_{ijt} + \beta_2 Death_{ijt} \times AllStar_i + \beta_3 Death_{ijt} \times LoneWolf_i + \\ \beta_4 Death_{ijt} \times Maven_i + \gamma_{it} + \mu_{jt} + \delta_t + \phi_{ij} + \varepsilon_{ijt} \quad (2.2) \end{aligned}$$

where the new dependent variable $\tilde{Y}_{-ijt} = Y_{-ijt}$ if $Y_{-ijt} \geq 1$, and $\tilde{Y}_{-ijt} = 1$ if $Y_{-ijt} = 0$. To distinguish between the values of $\tilde{Y}_{-ijt}$ where $Y_{-ijt} = 1$ and where $Y_{-ijt} = 0$, I create a variable d_{ijt} where $d_{ijt} = 1$ if $Y_{-ijt} = 0$ and add it as an independent variable. This technique allows for straightforward interpretation of the coefficients as well as allowing the variables of interest to be interacted with continuous variables (Pakes and Griliches, 1980; Acemoglu and Linn, 2004).[14] In addition, by adopting a linearization of the functional form of the regression, I can carry out instrumental variable analysis through two-stage least squares estimation (2SLS), which is much more cumbersome in non-linear models such as the poisson outlined in Equation 2.1.

I present both specifications (Equations 2.1 and 2.2) across two main samples. The first sample only includes dyads in which a star dies. Twenty-eight of the 415 immunologists die, and they each have an average of 29.1 coauthors, resulting in a sample consisting of 816 dyads. The second sample includes all immunologist-coauthor dyads, regardless of whether the immunologist dies or not. The full set of 415 immunologists has, on average, 58.3 coauthors, resulting in a sample of 24,175 dyads. Across both modeling specifications, the regression analysis follows a differences-in-differences style estimation. For the sample consisting solely of dyads where the immunologist dies, I use variation in the death of immunologist i to estimate the relationship between death and the change

---

[14] The magnitude of interaction effects in nonlinear models do not equal their marginal effects and thus makes direct interpretation of results difficult (Ai and Norton, 2003).

in productivity of coauthor j. For the sample consisting of all dyads, I introduce a second dimension of variation, comparing the possible disparity between scientists whose coauthors die and those whose do not. Since only 7% of the 415 immunologists in the sample die (the "treated" group), the remaining 93% in the full sample serve as a de facto control group.

2.6 Results

2.6.1 Main Results

This study asks two main research questions. First, using the old dichotomous definition of star scientists, do productivity stars generate more spillovers than Non-Stars? Second, using my new taxonomy of star scientists, to what extent do different star types influence the productivity of others? The first question was already asked and partially answered in the very important work by Azoulay, Graff Zivin, and Wang (2008) (AGW). As mentioned earlier, however, the empirical results that AGW present, whereby a scientist's productivity decreases after the death of a coauthor, are unable to rule out two possibly conflating effects. First, because AGW condition their sample on the top 5% of life scientists and solely examine the influence of their deaths on the productivity of their coauthors, they are unable to examine the effect of spillovers on non-stars, as non-stars are omitted from their sample. They do show evidence that as a scientist's citations increase so do his spillovers, but this still is conditioned on being in the top 5%. Second and similarly, without the inclusion of different star types in their sample, they are unable to disentangle the decrease in productivity of coauthors due to the loss of an intellectual link rather than the decrease in productivity due to the disruption caused by the death of a former colleague.

Before I explore the extent to which different star types influence the productivity of others, I feel it prudent to address the question of the extent to which stars generate more spillovers than non-stars, first by replicating AGW's results and then by including controls for non-stars to appropriately deal with the alternate explanations outlined above.

Table 2.6 presents the main results of Equation 2.1 from section 2.5. To allow for readily comparable coefficients to those presented by AGW, the sample used in Table 2.6 only includes dyads where the immunologist dies. Specification 1 further restricts the sample by only including immunologists who have high productivity, that is, All-Stars and Lone Wolves. This definition and sample are identical to those used in AGW. The coefficient on death is -0.173, which translates into a 15.9%[15] decrease in productivity, somewhat larger than AGW's range of estimates between -5 and -10%. Specification 2 returns the sample to all dyads with a death. The death variable is interacted both with immunologists who have high (All-Stars and Lone Wolves) and average (Mavens and Non-Stars) productivity. The coefficient on the high productivity interaction, significant at the 5% level, indicates that the productivity of coauthors of high productivity stars decreases by 17.4% (-0.191), again somewhat larger than the findings of AGW. The coefficient on average productivity immunologists is not statistically distinct from 0, indicating that average productivity immunology stars in the aggregate have little influence on the productivity of their coauthors. Specification 3 is identical to Specification 2, but instead of interpreting the coefficients on the influence of high productivity on coauthor productivity relative to when the immunologist was alive, the omitted category in Specification 3 is the death of an average productivity immunologist. Interestingly, the null hypothesis that the productivity effects of the death of a high productivity star is different from the death of a average productivity star cannot be rejected. Specification 4 looks at the average effect of death on a coauthor's productivity for all star types. That is, the average effect of the death of an immunologist in the sample decreases the productivity of their coauthors by 4.4%. This value, however, is highly insignificant.

Specification 5 introduces the second dimension of the star taxonomy: helpfulness.

---

[15] $exp(-0.173) - 1 = -0.159$

The omitted category for Specification 5 is the death of a Non-Star, and so all coefficients should be interpreted as relative to the productivity effects of the death of a Non-Star. Most strikingly, the death of an All-Star decreases the productivity of his coauthors by 38.6% (-0.488), the death of a Lone Wolf decreases the productivity of his coauthors by 23.6% (-0.269), and the death of a Maven decreases the productivity of his coauthors by 38.7% (-0.489). All coefficients are significantly different from 0 at the 10% level, while the Maven coefficient is significant at the 1% level.

Table 2.7 continues to present the main results of Equation 2.1 but differs from the results presented in Table 2.6 in that I estimate the full sample (dyads where the focal immunologist dies and dyads where the focal immunologist does not die). The models run in Table 2.7 are identical to those run in Table 2.6 apart from the different sample. As in the previous table, the sample in Specification 1 only includes high productivity immunologists (All-Stars and Lone Wolves). The average effect of the death of a high productivity star is not only quantitatively smaller than in the sample with only dyads where a death occurs but also statistically insignificant. Specification 2 returns to the full sample and examines the effect of death on coauthor productivity by whether or not the immunologist has high or average productivity. The death of a high productivity star results in a 31.7% (-0.381) decrease of his coauthor's productivity and is highly significant. Interestingly, the death of an average productivity immunologist also has a negative impact on his coauthor's productivity, decreasing the subsequent quality adjusted output by 17.1% (-0.188). This indicates that when thinking about stars on a productivity continuum, the death of a coauthor, irrespective of his productivity, negatively influences the productivity of his coauthors, albeit with different magnitudes. Specification 3 changes the omitted variable to the death of an average productivity immunologist, allowing for the direct test of the null hypothesis that high productivity immunologists have a larger impact on the productivity of their coauthors (they generate more spillovers) than average productivity immunologists. While the coefficient indicates that the productivity of coauthors of high productivity immunologists who die is 17.6% lower than average productivity immunologists who die, this value is not statistically distinct from 0. Specification 4 makes no distinction between immunologist star types and shows that the average decrease in productivity of coauthors of immunologists who have died is 23.2% (-0.264).

Specification 5 introduces the new taxonomy star types. In this specification, death has an average 11% negative effect on the productivity of coauthors but is statistically insignificant. The change in a coauthor's productivity after a death varies by the type of star that dies. Coauthors of All-Stars are 40.0% (-0.510) less productive, and coauthors of Mavens are 39.0% (-0.494) less productive. The coauthors of Lone Wolves are 21.4% (-0.241) less productive, but this value is not statistically distinct from 0. All of these coefficients are relative to the decrease in productivity of coauthors of a Non-Star who dies. The variables that appear in Specification 5 are the main variables of interest for this study, and their relationships will be shown in various functional forms throughout this paper. The Lone Wolf and Maven coefficients are statistically distinct from one another at the 5% level.

To compare these results with those of AGW, recall that their range in productivity decreases due to a coauthor's death is between -5 and -10%. As can be seen from Tables 2.6 and 2.7, the estimates I present are at times outside of this bound. One of the reasons for this difference may lie in the change of setting. Where AGW look at a range of life science disciplines, this study only examines immunologists. Second, AGW construct a control group using a form of propensity score matching to find two "nearest neighbor" control stars for every star who dies. My study, instead, uses all immunologists who do not die in my sample of 415 immunologists to form a control group. If either the immunologists used in the control group or their coauthors are less likely to generate spillovers (for a myriad of factors) than the immunologists who die or their coauthors, then my estimates will appear larger in magnitude. As can be seen from Table 2.5, the coauthors of immunologists who die do have higher average publication, citation, and Impact Factor-weighted publication counts than the coauthors of immunologists who do not die. This, however, does not directly imply that there should be any difference in the spillover levels across these two groups. Regardless, the range of parameter estimates do appear similar enough to assuage a reader's concern that AGW and this study are examining different phenomenon.

Overall, the significance of the star types from the new taxonomy appear both significant and stable. It does appear that Mavens are different from Lone Wolves. Furthermore, high productivity immunologists – All-Stars and Lone Wolves – each have different effects on the productivity of their coauthors when they die. The next section examines the robustness of these relationships.

2.6.2 Robustness Checks

Table 2.9 presents OLS estimates of Equation 2.2 from Section 2.5. Specifications 1 and 2 replicate the results from Specification 5 from Tables 2.6 and 2.7, where a coauthor's change in productivity is a function of the type of star who died. The coefficients from this linear model are qualitatively similar to those presented in the poisson fixed effects models estimated by quasi maximum likelihood (QML), and as such, I feel confident using the OLS to estimate the relationship despite the count nature of the dependent variable. Specification 2, which makes use of the entire sample, shows that the death of an All-Star decreases the productivity of his coauthors by 24% (-0.274), while the death of a Maven decreases the productivity of his coauthors by 21.6% (-0.243). Both of these estimates are significant at the 1% level. Again, as in previous specifications, the effect of the death of a Lone Wolf is statistically indistinguishable from the effect the death of a Non-Star has on his coauthors.

Of the 415 immunologists in the sample, 28 die. While my identification of the impact of their deaths comes from the variation in productivity of their coauthors, concerns of outliers driving the results may still exist. In Specifications 3 and 4, instead of looking at the death of the sample of immunologists, I examine their "exit" decisions. An exit is defined as the year in which an immunologist stops publishing for at least a four-year period. According to academic immunologists, if an immunologist has failed to publish a single manuscript in four years, then it is fairly reasonable to assume that this person has exited the risk set of publishing. An exit can occur for a number of reasons: retirement, decrease in productivity, move to industry, and of course death. Using the full sample in Specification 4, the results differ somewhat from the other tables. The exit of an All-Star is both economically and statistically insignificant, and the decrease in productivity of coauthors of Lone Wolves and Mavens are not statistically distinct from one another. Three major explanations for these results exist. First, All-Stars may announce their exits well in advance, so that coauthors have ample time to adjust to the anticipated loss of the All-Star. Second, if an All-Star exits, he may still be active in an advising capacity, thus mitigating any losses that may have befallen his coauthors. Third, coauthors of exited All-Stars aren't very productive before or after the exit, and thus no decline is observed. These alternate reasons are unknown to the econometrician, and so I must search for some form of exogenous variation.

To remedy the endogeneity of the exit decision, Specifications 5 and 6 report results from an instrumental variable (IV) two stage least squares (2SLS) estimation where the death measure acts as an instrument for the exit decision. While results presented indicate a relationship between the death of an immunologist and the productivity of his coauthors, the relationship is driven entirely through the exit of the immunologist from the coauthor's coauthorship network. As such, the death of an immunologist should not affect his coauthor's productivity in any way other than through his exit, making death an appropriate candidate for an instrument of exit. I present linear probability models of the first stage estimates whereby the likelihood of an exit is a function of death in Table 2.8. As can be seen from the first stage estimates in Table 2.8, the death of an immunologist has a highly significant impact on the likelihood of the immunologist exiting the sample, almost by definition. Having an appropriate instrument, however, increases the level of variation and attenuates the likelihood that outliers are driving the results. The IV estimates presented in Specification 6 from Table 2.9 reveal that the exit of an All-Star decreases the productivity of his coauthors by 42.1% (-0.546), while the exit of a Maven decreases the productivity of his coauthors by 35.9% (-0.445). Both of these coefficients are statistically significant at the 5% level, while the effect of a Lone Wolf's exit on his coauthor's productivity is statistically insignificant.

Table 2.10 replicates Table 2.9 but changes the dependent variable from the count of Impact Factor-weighted publications to the count of citation weighted publications. As in Table 2.9, I provide OLS estimates for both the effect of a death and exit on a coauthor's productivity in addition to instrumenting exit with death. With the change of the dependent variable, the results are somewhat different. The death of a Maven has a larger effect on the decrease in productivity of his coauthors than the death of an All-Star, although these coefficients are not statistically different from one another. More interesting, however, is that for the first time the exit of a Lone Wolf has a larger negative effect on the productivity of his coauthors than the exit of an All-Star, even after instrumenting for exit with death. The effect of a Maven's exit still has the largest negative effect on his coauthors, but the coefficient is not statistically distinct from the Lone Wolf coefficient. While all three major star types – All-Stars, Lone Wolves, and Mavens – negatively affect the productivity of their coauthors with respect to citation-weighted publications, no one star type has a larger effect that is statistically significant. One of the main reasons for this finding may be the idiosyncratic way in which citations are awarded and the effect an individual has on generating citations. While a lone individual can greatly influence the quality of a manuscript and thus the quality of the journal it is accepted by, the process by which an article collects citations is less understood.

Table 2.11 presents results from estimating my main regression specification of interest by means of a fixed effects negative binomial. While the fixed effects poisson estimator with robust standard errors is much preferred over the negative binomial estimator (as the poisson estimator presents more conservative [larger] standard errors), the negative binomial estimator is still quite heavily used in strategy, sociology, and economics in the application of count models and as such is included to ensure that the results still look reasonable. Specification 2 shows results from the full sample and finds results consistent with other estimators, whereby the death of an All-Star or Maven decreases the productivity of his coauthors in excess of 20%.

Table 2.12 moves away from the discrete definition of productivity and helpfulness and instead looks at the continuous relationship between the effect of an exit or death interacted with continuous measures of productivity and helpfulness. I have converted the measures of productivity and helpfulness to logs so that the coefficients may be interpreted as elasticities, which allows for a unit-free interpretation of the parameters. Specification 1 shows the effect of an immunologist's exit on the productivity of his coauthors. A doubling in the productivity of the immunologist, where productivity is measured by citations received, results in a 5.2% decrease in the productivity of his coauthors. Helpfulness has no statistical effect on the productivity of coauthors when an immunologist exits. Specification 2 looks at deaths instead of exits. A doubling in the helpfulness of an immunologist who dies reduces the productivity of his coauthor by 13.4%, a much higher level than in the case of exits, again raising the need for an IV estimation to address the endogeneity of exits. Specification 3 instruments exits with death and finds that a doubling of the helpfulness of an immunologist is associated with an 18.7% decrease in productivity of his coauthors, while changes in productivity have no effect on the coauthor's productivity. I change the dependent variable to citation-weighted publication counts in Specifications 3 through 6. Immunologist productivity has a statistically significant effect on the productivity of coauthors when an immunologist exits, but looking at deaths and exits instrumented by death, the effect of helpfulness strengthens yet retains only marginal statistical significance.

2.7 Alternate Explanations

At least three alternate explanations may account for the relationship observed between the death of various star types and the subsequent decrease in performance of their coauthors. First, the decrease in a coauthor's performance may come from a decrease in funding that had been provided by the now deceased immunologist. Second, the status of the immunologist is artificially increasing the perceived performance of his coauthors and thus, after his death, the coauthors return to their natural steady-state. This casts doubt on the claim that the decrease in performance of a coauthor when an immunologist dies is due to the elimination of spillovers. Third, the effects observed are due to the influence of institution- and university-specific factors that influence the productivity of coauthors.

The first alternate consideration for the presence of the report results comes from the concern that an immunologist's funding largely drives coauthor productivity. The concern is that because funding is almost always linked to a primary investigator, funding would stop with an immunologist's death, and so the relationship we witness whereby a coauthor's performance decreases with the death of an immunologist is simply being driven by the omitted funding variable. For this to be a viable alternate explanation, however, I must assume that a coauthor is able to benefit or make use of the immunologist's funding yet not include the immunologist as a coauthor. The probability of such an occurrence is very low as any use of funds will surely be tied to a coauthoring arrangement. Furthermore, recall that the dependent variable used throughout is the quality adjusted count of papers written by the coauthor without the immunologist. If the coauthor is benefiting from the immunologist's funding and consequently frequently coauthoring with the immunologist, then the coauthor's pre-death publications with the immunologist will not be counted. The empirical exercise of this paper is to compare pre- and post-death publishing rates of coauthors. If the pre-death rates are lower due to frequent coauthorship with the focal immunologist (and thus netted out), then the observed change in productivity after the immunologist dies should be negligible, thus biasing my results in an opposite direction from what is observed.

The second alternate explanation for the reported results is the conflation of performance with status. The concern is that a coauthor experiences positive performance due to an association with a high status immunologist (Merton, 1973). For this to be true, status cannot act as an information signal, wherein due to information asymmetry association with an immunologist conveys quality onto the coauthor, consequently increasing his performance. If in this context status acts as a quality signal, then the signal should not be weakened once the immunologist dies, and consequently a decrease in a coauthor's performance after the death of an immunologist is unlikely to be associated with status effects. Furthermore, if status is driving the results reported, then I would not necessarily predict the strong effect of the death of a Maven. Moreover, if status is driving my results, then one would expect the effect of a Lone Wolf's death to have even larger negative effects on their coauthor's productivity, as All-Stars and Lone Wolves are considered high status individuals.

Third, heterogeneity in resources available to institutions or universities may increase the productivity of immunologists. While this is certainly true, the only way for institutional effects to influence the performance of coauthors is if at the time of an immunologist's death the coauthor changes institutions. I employ dyad fixed effects across all specifications, and so any time-invariant characteristics that do not alter over the course of the panel, such as institutional setting, will be captured by the fixed effects. Lastly, it is very unlikely that a coauthor changes institutions in response to a former colleague's death, and thus institutional effects should have no influence on the results observed.

2.8 Discussion and Conslusion

This paper develops a new taxonomy of star scientists for the purpose of identifying scientists most likely to generate spillovers. I expand the current conceptualization of star scientists, which presently only examines an individual's productivity, by adding a social behavior dimension: helpfulness. Helpfulness is the extent to which an individual is beneficial to others. By dividing immunologists into four classifications along the dimensions of productivity and helpfulness, I examine the magnitude of the decrease in performance of coauthors of immunologists who have died. I define All-Stars as scientists with both high productivity and helpfulness, where high is defined as being in the top 5% of the distribution. Lone Wolves are scientists with high productivity but average helpfulness. Mavens are scientists with average productivity but high helpfulness. Non-Stars make up the fourth classification and are average in productivity and helpfulness. Mavens and All-Stars have the largest negative impact on the publishing rates of their coauthors when they die, indicating the loss of a source of a spillover. The death of an All-Star on average decreases the performance of his coauthors by 35%, while the death of a Maven decreases the performance of his coauthors by 30%. These findings are robust to a series of controls and specifications.

These findings have several important implications for both theory and practice. First, the resource-picking mechanism in the resource-based view of the firm requires somewhat imperfect information to outsmart the resource market. This is difficult to do with public and highly verifiable resources such as a high productivity star. High helpfulness stars, on the other hand, are more difficult to measure, resulting in possible information asymmetries and a possible source of sustainable competitive advantage. Second, the learning by hiring literature (Song, Almeida, and Wu, 2003) has mostly focused on the acquisition of knowledge through the hiring of individuals. Yet, as has been argued throughout this paper, if factor markets are efficient, then a recruited engineer, for example, should be able to capture his full output, which includes his embodied knowledge, through his salary. A logical extension to this literature is to think about the extent to which different star types affect a firm's ability to learn through the hiring process. Third, alongside AGW, this paper is one of the first studies to both measure and find conclusive support of human capital spillovers from star scientists. By focusing on scientists most likely to generate spillovers through the formation of a new taxonomy, I find large economic and statistically significant spillovers.

Strategic hiring decisions are no longer the sole domain of human resources but rather require top corporate-level decision making as human capital has a large impact on firm performance. For many knowledge-based companies, the largest expense is the cost of human capital in the form of salaries. Developing an effective organizational design to properly manage the main inputs of a firm's innovation production function is of great importance.

From a public policy standpoint, a more nuanced understanding of human capital will surely inform the debate on regional clusters. One of the main benefits of clusters is both the generation of and the ability to absorb non-rival knowledge spillovers. Policy makers go to great lengths in developing appropriate incentives to optimize cluster structure with the goal of maximizing welfare. Understanding what types of human capital are most likely to generate knowledge spillovers is of critical concern.

While the development of this new taxonomy provides a tool with which to examine human capital along a second dimension, its construction raises additional questions. How are Mavens priced relative to Lone Wolves? Does the market appropriately price spillovers? The mechanism presented in this study is that of a strong tie formation. How then do All-Stars and Mavens diffuse knowledge and spillovers across weak ties? These questions are left to future research.

A number of limitations with this study still exist. First is the endogenous nature of acknowledgements. No clear or externally enforced rule exists over the administration of acknowledgements. Consequently, acknowledgements may be strategically applied for personal gains, such as winning favor with a journal referee or editor. However, if acknowledgements are bestowed upon individuals who are not helpful but instead in a position of authority, then we would expect additional noise to enter the helpfulness measure and consequently bias results towards 0 and in the opposite direction from the results presented here.

Second, the main mechanism of spillovers comes from the establishment of a social relationship through the formation of a coauthoring relationship. Clearly, this is not the only mechanism by which spillovers transfer nor do I purport this to be so. This mechanism should be viewed within the context of a larger research agenda that explores the ways in which different star types generate spillovers to others.

Lastly, the external validity of this setting may be limited. While the objective functions of firms and academic departments are not entirely orthogonal, they are certainly different. While a firm setting would be ideal, the difficulty in obtaining both individual level productivity and helpfulness measures is greatly encumbering. Nonetheless three factors contribute to the attractiveness of studying immunologists. First, because acknowledgement patterns vary across disciplines, it is important to isolate this heterogeneity by focusing on a single distinct discipline. Second, because of this heterogeneity in acknowledgement norms found across academia, it is important to look at a discipline where acknowledgments are both present and applicable to my construct of helpfulness. And third, measures on helpfulness are incredibly hard to identify, let alone access for a cross section of firms, and as such are not the appropriate focus of a study such as this. A survey instrument, however, may be able to capture individual levels of helpfulness in a firm setting. This avenue is left open for future research.

This study presents preliminary evidence of the productivity gains associated with coauthoring with helpful scientists. In doing so, it makes three important contributions. First, it extends the current dichotomous conceptualization of star scientists by explicitly defining star classification not only along the dimension of productivity but also the spectrum of helpfulness, thus developing a new taxonomy of star scientists. Second, it provides a measure by which helpfulness can be empirically tested: acknowledgements. Third, it attempts to establish a causal link between coauthoring with All-Stars and Mavens and an increase in productivity.

The traditional method of bundling together All-Stars and Lone Wolves is quite problematic. All-Stars and Lone Wolves are quite different in their spillover generating capabilities. Furthermore, Mavens, who under the current dichotomous conceptualization of star scientists are classified as Non-Stars, are actually quite important in generating spillovers. As such, our current classification of star scientists has possibly been systematically overvaluing Lone Wolves while undervaluing Mavens.

Figure 2.1: Productivity and Helpfulness Distribution: N=415

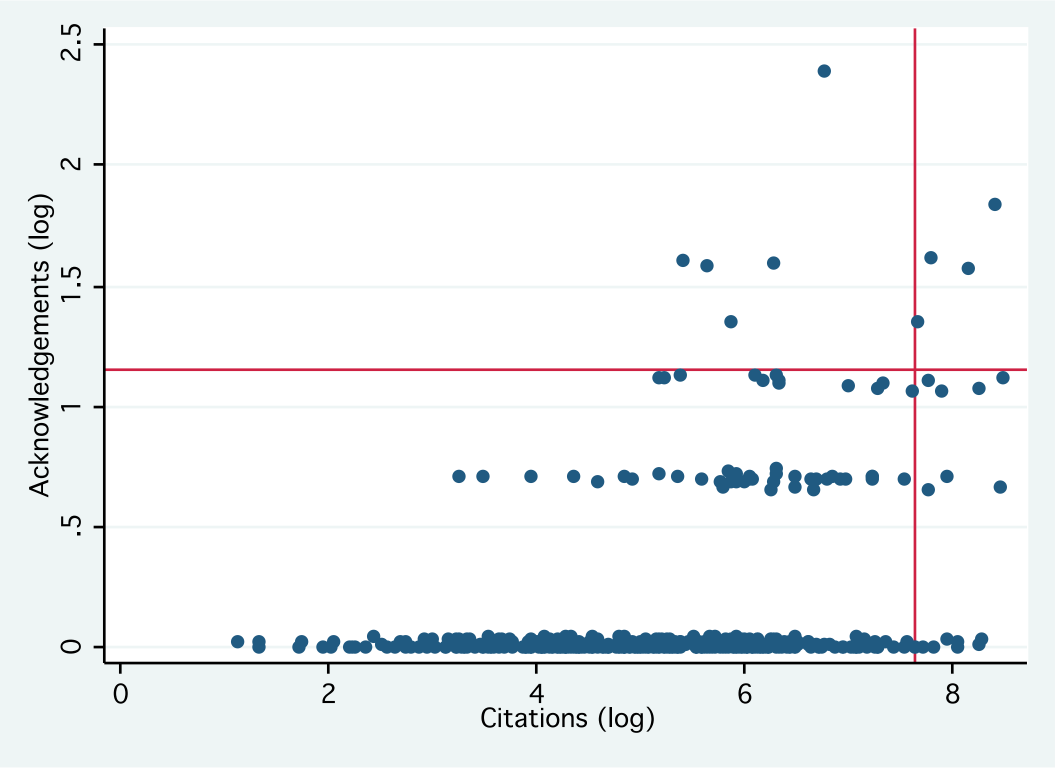

Table 2.1: A New Taxonomy for Star Scientists

	Average Productivity	High Productivity
High Helpfulness	Maven	All-Star
Average Helpfulness	Non-Star	Lone Wolf

Table 2.2: Star Scientist Classifications - Top 5%

	Average Productivity	High Productivity
High Helpfulness	**Maven** Total N = 5 Dyads = 189 Average coauthors = 37.8 Died N = 3 Died Dyads = 68 Average coauthors = 22.66	**All-Star** N = 4 Dyads = 188 Average coauthors = 47 Died = 2 Died Dyads = 22 Average coauthors = 11
Average Helpfulness	**Non-Star** Total N = 390 Dyads = 23,125 Average coauthors = 59.29 Died N = 18 Died Dyads = 490 Average coauthors = 27.22	**Lone Wolf** N = 16 Dyads = 673 Average coauthors = 42.06 Died = 5 Died Dyads = 236 Average coauthors = 47.2

Table 2.3: Variable Descriptive Statistics. Full Sample: N = 575,483

	Variable	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11	12
1	Impact Factor Pubs	9.737	18.43												
2	Ln Impact Factor Pubs	1.376	1.34	0.74											
3	Citation-Weighted Pubs	22.239	56.10	0.20	0.33										
4	Death	0.012	0.11	0.02	0.04	−0.01									
5	All-Star Death	2.E-4	0.01	0.01	0.01	0.00	0.13								
6	Lone Wolf Death	0.003	0.06	0.02	0.03	0.00	0.55	0.00							
7	Maven Death	0.001	0.02	0.01	0.01	0.00	0.22	0.00	0.00						
8	Non-Star Death	0.007	0.08	0.01	0.02	−0.01	0.79	0.00	−0.01	0.00					
9	Exit	0.085	0.28	0.07	0.11	−0.03	0.23	0.05	0.11	0.08	0.18				
10	All-Star Exit	0.001	0.04	0.02	0.02	0.00	0.05	0.40	0.00	0.00	0.00	0.12			
11	Lone Wolf Exit	0.007	0.08	0.02	0.04	0.00	0.23	0.00	0.42	0.00	−0.01	0.28	0.00		
12	Maven Exit	0.002	0.05	0.01	0.02	0.00	0.10	0.00	0.00	0.47	0.00	0.16	0.00	0.00	
13	Non-Star Exit	0.074	0.26	0.06	0.10	−0.03	0.15	0.00	−0.02	−0.01	0.20	0.93	−0.01	−0.02	−0.01

Table 2.4: Variable Descriptive Statistics. Died Dyads Sample: N = 25,968

	Variable	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11	12
1	Impact Factor Pubs	11.029	17.392												
2	Ln Impact Factor Pubs	1.569	1.348	0.78											
3	Citation-Weighted Pubs	25.142	55.955	0.23	0.33										
4	Death	0.255	0.436	0.07	0.11	−0.09									
5	All-Star Death	0.005	0.068	0.03	0.03	−0.01	0.12								
6	Lone Wolf Death	0.077	0.267	0.07	0.10	−0.02	0.49	−0.02							
7	Maven Death	0.012	0.110	0.02	0.04	−0.03	0.19	−0.01	−0.03						
8	Non-Star Death	0.161	0.367	0.02	0.03	−0.08	0.75	−0.03	−0.13	−0.05					
9	Exit	0.255	0.436	0.13	0.18	−0.06	0.57	0.12	0.24	0.19	0.43				
10	All-Star Exit	0.014	0.117	0.07	0.06	0.00	0.02	0.58	−0.03	−0.01	−0.05	0.20			
11	Lone Wolf Exit	0.056	0.229	0.07	0.11	−0.01	0.33	−0.02	0.70	−0.03	−0.11	0.41	−0.03		
12	Maven Exit	0.042	0.201	0.05	0.07	−0.02	0.02	−0.01	−0.06	0.53	−0.09	0.36	−0.02	−0.05	
13	Non-Star Exit	0.144	0.351	0.07	0.09	−0.06	0.48	−0.03	−0.12	−0.05	0.67	0.70	−0.05	−0.10	−0.09

Table 2.5: Coauthor Means (Std. Dev.) by Dyad Type

	Papers	Citations	Impact Factor Papers	Citation Weighted Papers
Full Sample				
All-Star	2.43	94.65	11.37	28.67
N = 4,659	(4.07)	(206.4)	(20.27)	(65.1)
Lone Wolf	2.39	96.04	11.31	30.27
N = 18,908	(4.09)	(209.7)	(18.92)	(65.53)
Maven	2.61	105.70	11.18	30.43
N = 5,285	(3.48)	(209.07)	(15.54)	(63.8)
Non-Star	2.33	71.19	9.65	21.83
N = 546,631	(4.2)	(172.93)	(18.42)	(55.55)
Died = 1				
All-Star	3.16	139.47	15.86	35.80
N = 792	(4.13)	(281.01)	(23.48)	(107.65)
Lone Wolf	2.83	103.65	12.24	28.82
N = 7,879	(3.94)	(215.39)	(18.61)	(62.2)
Maven	3.44	124.25	12.91	27.62
N = 2,551	(4.28)	(235.42)	(17.31)	(48.59)
Non-Star	2.44	74.58	9.80	22.18
N = 14,746	(3.48)	(163.75)	(16.19)	(48.92)
Died = 0				
All-Star	2.28	85.47	10.46	27.21
N = 3,867	(4.05)	(186.21)	(19.42)	(52.18)
Lone Wolf	2.08	90.60	10.65	31.32
N = 11,029	(4.17)	(205.38)	(19.11)	(67.79)
Maven	1.84	88.38	9.57	33.05
N = 2,734	(2.25)	(179.38)	(13.49)	(75.18)
Non-Star	2.33	71.09	9.65	21.82
N = 531,885	(4.22)	(173.17)	(18.48)	(55.72)

Table 2.6: Poisson QML Baseline Model - Dyads with a Death

Dependent Variable:	*Coauthor Impact Factor Weighted Publication Counts*				
	(1)	(2)	(3)	(4)	(5)
Death	-0.173^{+} (0.092)		0.036 (0.098)	-0.044 (0.076)	0.098 (0.107)
Death X High Productivity		-0.191^{*} (0.094)	-0.227 (0.139)		
Death X Average Productivity		0.036 (0.098)			
All-Star Death					-0.488^{+} (0.257)
Lone Wolf Death					-0.269^{+} (0.143)
Maven Death					-0.489^{**} (0.130)
Dyad Fixed Effects	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓
Star Age Cohort FE	✓	✓	✓	✓	✓
Coauthor Age Cohort FE	✓	✓	✓	✓	✓
Observations	8671	25968	25968	25968	25968
Number of Dyads	258	816	816	816	816
Log Likelihood	-55362	-151631	-151631	-151830	-151225

QML robust star cluster adjusted standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.7: Poisson QML Baseline Model - Full Sample

Dependent Variable:	*Coauthor Impact Factor Weighted Publication Counts*				
	(1)	(2)	(3)	(4)	(5)
Death	-0.076 (0.112)		-0.188^{+} (0.110)	-0.264^{**} (0.084)	-0.115 (0.117)
Death X High Productivity		-0.381^{**} (0.125)	-0.194 (0.160)		
Death X Average Productivity		-0.188^{+} (0.110)			
All-Star Death					-0.510^{+} (0.271)
Lone Wolf Death					-0.241 (0.165)
Maven Death					-0.494^{**} (0.119)
Dyad Fixed Effects	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓
Star Age Cohort FE	✓	✓	✓	✓	✓
Coauthor Age Cohort FE	✓	✓	✓	✓	✓
Observations	23567	575483	575483	575483	575483
Number of Dyads	861	24175	24175	24175	24175
Log Likelihood	-147400	-3055558	-3055558	-3055758	-3055102

QML robust star cluster adjusted standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.8: OLS - Linear Probability Model - First Stage Estimates of Exit

Dependent Variable:	Exit	All-Star Exit	Lone Wolf Exit	Maven Exit	Non-Star Exit
	(1)	(2)	(3)	(4)	(5)
Death	0.429**				
	(0.082)				
All-Star Death		0.557**			
		(0.100)			
Lone Wolf Death			0.488**		
			(0.161)		
Maven Death				0.613**	
				(0.039)	
Non-Star Death					0.530**
					(0.075)
Constant	0.388**	-0.007	0.023	0.002	0.361**
	(0.077)	(0.006)	(0.018)	(0.005)	(0.075)
Dyad Fixed Effects	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓
Star Age Cohort FE	✓	✓	✓	✓	✓
Coauthor Age Cohort FE	✓	✓	✓	✓	✓
Observations	575483	575483	575483	575483	575483
Number of Dyads	24175	24175	24175	24175	24175
Log Likelihood	180066	1245618	822523	1114320	217141
R^2	0.39	0.07	0.15	0.13	0.39

Star cluster adjusted standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.9: OLS and 2SLS Model with Deaths and Exits I

Dependent Variable:	*Log of Coauthor Impact Factor Weighted Publication Counts*					
Dyads Included	Death	All	Death	All	Death	All
Estimation	OLS	OLS	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Death	-0.035	-0.091				
	(0.045)	(0.061)				
All-Star Death	-0.223^{*}	-0.274^{**}				
	(0.103)	(0.097)				
Lone Wolf Death	-0.082	-0.090				
	(0.067)	(0.069)				
Maven Death	-0.192^{*}	-0.243^{**}				
	(0.088)	(0.077)				
Exit			-0.006	-0.148^{**}	-0.240^{+}	-0.194^{+}
			(0.066)	(0.024)	(0.124)	(0.108)
All-Star Exit			0.007	-0.005	-0.341	-0.546^{*}
			(0.077)	(0.089)	(0.285)	(0.260)
Lone Wolf Exit			-0.132	-0.140^{**}	-0.174	-0.256
			(0.087)	(0.043)	(0.130)	(0.158)
Maven Exit			-0.171	-0.189^{+}	-0.434^{**}	-0.445^{**}
			(0.115)	(0.103)	(0.145)	(0.083)
Constant	1.078	0.744^{**}	1.003	0.804^{**}		
	(0.660)	(0.144)	(0.664)	(0.144)		
Dyad Fixed Effects	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓
Star Age Cohort FE	✓	✓	✓	✓	✓	✓
Coauthor Age Cohort FE	✓	✓	✓	✓	✓	✓
Observations	25968	575483	25968	575483	25968	575483
Number of Dyads	816	24175	816	24175	816	24175
Log Likelihood	-29755	-636510	-29753	-636045	-29928	-636336
Adjusted R^2	0.52	0.53	0.52	0.53	0.50	0.51

Star cluster adjusted standard errors in parentheses.
For Specifications 5 and 6, death instruments for Exit.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.10: OLS Model with Deaths and Exits II

Dependent Variable:	*Log of Coauthor Citation Weighted Publication Counts*					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV
Death	0.018	-0.002				
	(0.042)	(0.032)				
All-Star Death	-0.092	-0.096^{*}				
	(0.109)	(0.045)				
Lone Wolf Death	-0.103^{+}	-0.091				
	(0.060)	(0.068)				
Maven Death	-0.186	-0.179^{+}				
	(0.111)	(0.093)				
Exit			-0.058	-0.159^{**}	-0.065	-0.011
			(0.059)	(0.023)	(0.132)	(0.062)
All-Star Exit			-0.039	-0.019	-0.107	-0.176^{**}
			(0.081)	(0.054)	(0.168)	(0.050)
Lone Wolf Exit			-0.228^{**}	-0.083^{*}	-0.177^{*}	-0.186^{**}
			(0.069)	(0.036)	(0.076)	(0.071)
Maven Exit			-0.097^{+}	-0.086^{*}	-0.350^{+}	-0.292^{*}
			(0.049)	(0.037)	(0.207)	(0.133)
Constant	0.246	0.434^{*}	0.255	0.504^{**}		
	(0.616)	(0.177)	(0.657)	(0.176)		
Dyad Fixed Effects	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓
Star Age Cohort FE	✓	✓	✓	✓	✓	✓
Coauthor Age Cohort FE	✓	✓	✓	✓	✓	✓
Observations	25968	575483	25968	575483	25968	575483
Number of Dyads	816	24175	816	24175	816	24175
Log Likelihood	-32805	-718489	-32776	-718109	-32799	-718363
Adjusted R^2	0.66	0.67	0.66	0.67	0.65	0.65

Star cluster adjusted standard errors in parentheses.
For Specifications 5 and 6, death instruments for Exit.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.11: Robustness: Negative Binomial Fixed Effects

Dependent Variable:	*Coauthor Impact Factor Weighted Publication Counts*	
	(1)	(2)
Death	0.023	-0.052^{**}
	(0.024)	(0.019)
All-Star Death	-0.237^{*}	-0.260^{**}
	(0.095)	(0.091)
Lone Wolf Death	-0.090^{*}	-0.070^{*}
	(0.036)	(0.031)
Maven Death	-0.144^{*}	-0.224^{**}
	(0.058)	(0.057)
Constant	-1.712^{**}	-1.854^{**}
	(0.333)	(0.075)
Dyad Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Star Age Cohort FE	✓	✓
Coauthor Age Cohort FE	✓	✓
Observations	25968	575483
Number of Dyads	816	24175
Log Likelihood	-73415	-1477864

Standard errors in parentheses.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2.12: Robustness: Continuous Measures of Productivity and Helpfulness

Dependent Variable:	*Log of Coauthor Impact Factor Weighted Publication Counts*			*Log of Coauthor Citation Weighted Publication Counts*		
	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS	(6) IV
Exit	0.144*		-0.143	0.221**		0.363
	(0.062)		(0.263)	(0.077)		(0.230)
Exit X Helpfulness	-0.058		-0.187^{*}	0.017		-0.097
	(0.045)		(0.094)	(0.030)		(0.067)
Exit X Productivity	-0.052^{**}		-0.017	-0.069^{**}		-0.062^{*}
	(0.011)		(0.047)	(0.012)		(0.030)
Death		-0.074			0.198^{+}	
		(0.141)			(0.111)	
Death X Helpfulness		-0.134^{**}			-0.075^{+}	
		(0.046)			(0.045)	
Death X Productivity		-0.004			-0.033^{+}	
		(0.020)			(0.017)	
Constant	0.806**	0.741**		0.511**	0.433*	
	(0.146)	(0.144)		(0.176)	(0.177)	
Dyad Fixed Effects	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓
Star Age Cohort FE	✓	✓	✓	✓	✓	✓
Coauthor Age Cohort FE	✓	✓	✓	✓	✓	✓
Observations	575483	575483	575483	575483	575483	575483
Number of Dyads	24175	24175	24175	24175	24175	24175
Log Likelihood	-635901	-636513	-636222	-717926	-718487	-718294
Adjusted R^2	0.53	0.53	0.51	0.67	0.67	0.65

Star cluster adjusted standard errors in parentheses.
For Specifications 5 and 6, death instruments for Exit.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Chapter 3

The Price of All-Stars, Lone Wolves, and Mavens

3.1 Introduction

A large literature exists on the importance of human capital spillovers on firm learning (Song, Almeida, and Wu, 2003), knowledge flows between firms (Oettl and Agrawal, 2008), and economic growth (Jacobs, 1969; Lucas, 1988; Azariadis and Drazen, 1990). A separate but equally vibrant literature has emerged focusing on the impact of high-performing individuals – stars – on a host of economic outcomes, ranging from firm foundings (Zucker, Darby, and Brewer, 1998) to the firm-specificity of skill (Groysberg, Lee, and Nanda, 2008). Surprisingly, however, very little work has examined the human capital spillovers that arise from stars themselves. One of the first studies to examine these human capital spillovers in the context of high-performing individuals is the work of Azoulay, Graff Zivin, and Wang (2008) (AGW). In this study, AGW examine the change in performance of the coauthors of eminent scientists who die. They find that the deaths of eminent star scientists have a large negative impact on the performance of their coauthors, but they are unable to compare these externalities from eminent scientists with non-eminent scientists nor are they able to effectively determine if these performance externalities are priced.

Chapter 2 builds upon the limitations of the AGW study by developing a new taxonomy for star scientists, whereby I classify an individual is classified not only by his productivity but also by his helpfulness. In doing so, I am able to examine the effects of All-Stars (high productivity and high helpfulness), Lone Wolves (high productivity but average helpfulness), Mavens (average productivity but high helpfulness) and Non-Stars (average productivity and helpfulness) on the performance of their coauthors. The introduction of this new taxonomy reveals stark differences in the impact of different star types on the performance of their coauthors, whereby scientists with average productivity but high helpfulness (Mavens) have a larger positive impact on the performance of their coauthors than scientists with high productivity but only average helpfulness (Lone Wolves).

What Chapter 2 is unable to determine, however, is whether or not these positive performance effects that are provided to coauthors are actual spillovers; that is, are they unpriced benefits that occur outside of the market mechanism?

To understand why some types of human capital may not be properly priced and thus generate spillovers, I develop a theory of marginal productivity appropriation, whereby salary is positively associated with an individual's observable human capital and negatively associated with his unobservable human capital.[1] Because All-Stars and Lone Wolves are (by definition) high-productivity individuals, they are also more observable on the market and thus because of outside offers receive larger salaries due to matching offers at their current institutions (Lazear, 1986). The helpfulness of Mavens, however, is difficult to observe outside of an organization's boundaries, lowering their potential outside offers, resulting in their lower internal salaries. As such, I develop in this chapter

---

[1] Empirically I am unable to directly measure marginal productivity. Instead I must rely on an individual's total output to serve as a proxy for productivity. So while I am unable to get fully at the spillover question, the results reported later in this study are suggestive of mispricing.

hypotheses where All-Stars and Lone Wolves not only receive larger salaries than Non-Stars but also Mavens, despite the positive performance benefits of Mavens shown in Chapter 2.

To test these hypotheses, I collect a sample of all academic immunologists and economists within the University of California system. Due to annual salary disclosure legislation, the salaries of all University of California faculty are made available online. Findings corroborate my theoretical predictions: All-Stars, Lone Wolves, and Mavens all receive higher salaries than Non-Stars, but All-Stars and Lone Wolves receive higher salaries than Mavens. In addition, the more a scientist collaborates with colleagues on the same campus, the higher his salary. This positive effect increases for Mavens (stars who generate large spillovers), yet Lone Wolves (who generate significantly fewer spillovers than Mavens) receive larger salaries the more they collaborate with external scientists. As a whole, it appears that scientists are compensated largely for their productivity and not their helpfulness despite the positive performance benefits of helpfulness. As such, Mavens do indeed generate human capital spillovers.

This study is outlined as follows. The next section provides a brief overview of the literature on pricing human capital, which provides context for the hypotheses developed afterwards. Section 3.3 provides an outline of the data used in this study as well as the two empirical strategies employed. Section 3.4 presents the main results and robustness checks. Section 3.5 provides discussion of the results and concluding thoughts.

3.2 Literature and Theory

Firms and organizations are increasingly interested in strategically managing their “talent” (Guthridge, Komm, and Lawson, 2008). Rightfully so, as a small number of individuals generate a disproportionately large share of output. This finding holds for inventors (Narin and Breitzman, 1995; Ernst, Leptien, and Vitt, 2000), academic physicists (Lotka, 1926), financial security analysts (Groysberg, Lee, and Nanda, 2008), and even professional golfers (Brown, 2008). With respect to patenting inventors, the top 1% produce more than 15% of all patents issued by the United States Patent and Trademark Office (USPTO).[2] As evidenced by this skewed distribution, individuals who generate disproportionately large output – stars – can constitute important sources of an organization's output.

While stars themselves are highly productive, a tacit assumption within the literature has been that they also generate knowledge spillovers (Azoulay, Graff Zivin, and Wang, 2008). These human capital-induced knowledge spillovers have long been at the core of endogenous economic growth theory (Lucas, 1988; Romer, 1990) but still have received scant systematic empirical inquiry. Yet the importance of human capital spillovers for firm strategy is immense. Innovation, which is characterized by the recombination of known ideas (Nelson and Winter, 1982), is becoming increasingly difficult to conduct independently (Jones, 2009). As a result, increased collaboration is a necessary condition for continued knowledge growth and innovation (Wuchty, Jones, and Uzzi, 2007). Individuals, such as stars, who are characterized both by large personal productivity and by having a strong positive impact on the performance of others are clearly important agents to study for innovation and firm strategy.

The goal of the study presented in Chapter 2 is to test the tacit assumption that stars also produce externalities. Results reveal that solely focusing on an individual's personal productivity obfuscates the relationship between high productivity stars and their impact on the performance of others. Indeed, measuring stars along a single dimension (individual productivity) provides only crude insight into their impact on the performance of others. By including a second, social, dimension (helpfulness), I am able to build upon the current dichotomous star taxonomy (star versus non-star) and create a new taxonomy of star scientists that consists of four new star types. All-Stars are individuals who have

---

[2] Author's calculations.

both high productivity and helpfulness; Lone Wolves have high productivity but average helpfulness; Mavens have average productivity but high helpfulness; and Non-Stars have both average productivity and helpfulness. Using a sample of academic immunologists, I examine the impact of these four star types on the performance of their coauthors. To infer causality, I examine the change in performance of the coauthors of different star types when the stars die. A decrease in a coauthor's performance implies that the stars provide benefits to their coauthors prior to their deaths. I find that coauthors of All-Stars experience a 35% decrease in performance, coauthors of Mavens a 30% decrease in performance, and coauthors of Lone Wolves a 19% decrease in performance relative to the decrease in performance of the coauthors of Non-Stars.

These findings imply that stars do indeed positively impact the performance of their coauthors but that a scientist's helpfulness is a stronger predictor for these performance gains than his own personal productivity. The question that remains, of course, is whether or not these performance gains are spillovers. Without data on how stars are compensated, I have no means of determining whether or not the gains to the coauthors of All-Stars, Lone Wolves, and Mavens occur outside the realms of the market mechanism and thus truly constitute a spillover or externality. It is the goal of this study to determine the extent to which star scientists are compensated for the performance gains they generate for their coauthors.

3.2.1 Pricing Human Capital

The price or cost of an asset has direct bearing on an organization's profitability. Absent complementarities, a commoditized asset – an asset that is homogenous, non-appropriable, and perfectly tradable – cannot act as a source of sustainable competitive advantage (Peteraf, 1993). Suppose that this asset is human capital. If efficient factor markets exist, then individuals will be able to fully appropriate their value to an organization. That is, their marginal wage rate will be set to their marginal productivity

(Hirshleifer, Glazer, and Hirshleifer, 1998). As such, this individual cannot serve as a source of competitive advantage to the firm as any performance gains from employing the individual would be transferred back to him in the form of wages or alternate compensation. Work by Møen (2005) finds that any human capital gains that may arise from the mobility of technical personnel is largely internalized by the market.

However, should any of the aforementioned conditions fail, human capital may act as a source of competitive advantage. One such instance is when an individual's human capital is (at least partially) firm-specific (Becker, 1962). Firm-specificity of human capital by definition implies that a subset of an individual's skills only creates value at a specific firm. By extension, these skills are worthless at other firms, which imposes downward pressure on the individual's salary. As a result, a wedge is created between the individual's value to the firm and his wage rate. Because the value of this skill is firm specific, the firm is able to make use of an asset below its willingness to pay and thus can constitute a competitive advantage (Groysberg, Lee, and Nanda, 2008; Marx, Strumsky, and Fleming, 2009). While this scenario may apply to certain types of knowledge workers, it surely does not apply to scientists, whose human capital is largely general and not firm or institution specific.

In instances of no firm-specificity, market forces determine the price of human capital yet differences may still arise in the value of human capital to different institutions if asymmetry in observation of human capital exists. The portion of a scientist's human capital that is only observable within the boundaries of the firm should be compensated less than the portion of a scientist's human capital that is observable to the entire marketplace. Due to offer matching, the price of a scientist should increase with the observability of his human capital (Lazear, 1986).[3]

---

[3] Chapter 4 develops a simple model of scientist mobility where mobility is an increasing function of the scientist's observable human capital. The model makes a link between mobility and value to outside institutions, as the more observable a scientist's human capital is, the greater the likelihood of his receiving outside raiding offers, which increases the likelihood of mobility.

3.2.2 Knowledge Contracting

For the price of human capital to be close to its value in the absence of firm-specificity, the underlying human capital must be both observable and tradable (appropriable). This is true when we think of cases where human capital consists of the sum of an individual's schooling, years of experience, or, in the case of scientists, contents of their curriculum vitae. In these instances, individuals are able to verify their human capital (either through the showing of one's schooling transcripts or by reading a scientist's journal articles). Difficulties arise, however, when an individual's human capital consists largely of knowledge.

Knowledge is both unobservable (until it is revealed) and, absent property rights, difficult to trade, resulting in market failure (Arrow, 1962). No buyer of knowledge will agree to a price without observing the quality of the knowledge they are purchasing, yet no seller will disclose their knowledge, as once it is disclosed, his ability to command any fee has all but disappeared. In arms-length market transactions, we would observe this market failure, yet within the boundaries of a firm, this market failure can be averted (Kogut and Zander, 1992). If we view the decision of an employee to disclose information to the firm as a repeated game, then a firm should be willing to compensate an employee for said knowledge disclosure, for should the firm defect and not compensate, the cost of the loss in future knowledge disclosures by the employee could outweigh the cost of compensation. As such, firms are largely able to overcome the inherent market failures in the market for knowledge.

However, while individuals may disclose knowledge within the boundaries of an organization more readily, the observation of this knowledge disclosure is only observed within the boundaries of the organization. The knowledge disclosure discussed here is inappropriable largely because of a lack of paper trail, and so while the organization itself is aware of the knowledge disclosure, outside organizations are not. Consequently, because outside organizations are only able to observe the observable aspects of the individual's human capital and not the unobservable, the individual's market wage will be lower than if his unobservable skills are properly priced. Or, put differently, the value of the individual to his organization will exceed what the organization needs to pay him as his outside matching offers will be lower (Lazear, 1986). As a result, individuals with human capital that increases knowledge flows within organizations can act as sources of competitive advantage for the organization.

3.2.3 Pricing Star Scientists

If firms are able to pay individuals less than their marginal productivity when individuals possess firm-specific human capital and have difficult-to-observe human capital, then how do firms compensate stars? In particular, how do firms price star scientists?

Of course not all star scientists are equal. All-Star scientists are scientists with high helpfulness and productivity, while Mavens are scientists with high helpfulness but average productivity. Both of these stars have large positive impacts on the performance of their peers, in particular their coauthors. Conversely, Lone Wolves are scientists with average helpfulness but high productivity. They, like All-Stars, have high levels of personal productivity but make a smaller impact on the performance of their peers than All-Stars or Mavens (Chapter 2). Both productivity and helpfulness have positive performance implications for organizations. A firm derives value from human capital when the performance of the human capital exceeds its cost.[4] As such, understanding the relative costs/prices of these three star types is crucial for firm strategy, for if the characteristics of some of these stars predisposes them to less value appropriation, then they can be sources of firm competitive advantage.

All-Stars and Lone Wolves both have high (observable) productivity. Their productivity is readily observable in the marketplace, which increases their salaries. While in the short-run some of a scientist's productivity may be firm specific (labs, graduate students,

---

[4] Cost here includes opportunity costs.

etc.), in the long-run little of a scientist's productivity is overtly firm specific. As a result, both All-Stars and Lone Wolves should receive a higher salary than less productive Mavens and Non-Stars. Formally:

Hypothesis 1 (H1) All-Stars and Lone Wolves receive higher salaries than Mavens and Non-Stars.

Despite helpfulness providing performance gains to firms, this trait is less likely to be compensated as highly as the personal productivity of All-Stars or Mavens because it is not externally observable and as such does not generate internal matching offers from external offers (Lazear, 1986). Mavens do, however, command a salary premium above Non-Stars because of the compensation they receive from firms for their knowledge flows. As such, while Mavens will not be compensated as highly as more productive All-Stars and Lone Wolves, they will receive higher salaries than Non-Stars. Formally:

Hypothesis 2 (H2) Mavens receive higher salaries than Non-Stars.

3.2.4 Internal Collaboration

An organization's main objective with respect to publishing is to maximize total publication output. Organizations will not need to intervene in the structure of scientist collaborations if the incentives of scientists correspond with the objectives of the organization. Take the case of Lone Wolves, for example. Lone Wolves are scientists characterized by high personal productivity. By generating high quality publications, Lone Wolves not only benefit their organization but also themselves through higher salaries and (relatedly) greater mobility prospects (Chapter 4). As such, the issue of coauthor selection for a Lone Wolf largely surrounds which collaboration will result in the highest quality publication. As a result, Lone Wolves will seek the best coauthor from the universe of potential coauthors for a given project.[5] Since I define Lone Wolves as scientists who

---

[5] While this is clearly a two-sided matching problem, I am only focusing on one side for simplicity.

produce high quality publications, the probability of a Lone Wolf finding the optimal coauthor(s) within his own institution is low. As such, Lone Wolves are less likely to engage in internal collaboration than the average scientist.

Conversely, Mavens are scientists characterized by high helpfulness but only average personal productivity. The projects that Mavens engage in are of only average quality, and so the importance of coauthor matching for Mavens is less important than for Lone Wolves. To think about this differently, a high quality paper requires that the best scientists produce it. For a Lone Wolf, the number of potential coauthors (high personal productivity scientists) is small, and the probability that these potential coauthors are collocated with the Lone Wolf is low (although this non-clustering assertion is an assumption). Mavens, on the other hand, produce average quality manuscripts, and so the number of potential coauthors who can help them achieve an average quality manuscript is greater than the set of potential coauthors available to Lone Wolves. As such, conditioning on the quality of the manuscript, Mavens have more choice in coauthor selection than Lone Wolves, and by extension, there is a larger probability that a suitable coauthor for a Maven may exist within the boundaries of their own institution than the probability of a suitable coauthor for a Lone Wolf.

While a Maven's choice of coauthor will have a negligible impact on the quality of the publication produced (the Maven's incentive), if the Maven collaborates with an internal coauthor (a coauthor at the same institution as the Maven), then the institution benefits from the externalities produced by the Maven's helpful nature. As a result, a Maven's optimal collaborative form for an institution is to collaborate with an internal coauthor. Yet for the Maven, a slight preference may be given to external coauthors as more choice exists. If the gains to internal collaboration for the institution are larger than the gains to external collaboration for the Maven, then incentives can be created to encourage internal collaboration for Mavens. As a result, the effect of internal collaboration on salary should be heightened for Mavens compared to Non-Stars.

Hypothesis 3 (H3) Mavens who collaborate more internally receive higher salaries than Mavens who collaborate less internally.

3.3 Data and Methods

3.3.1 Sample

I construct two samples to test the hypotheses formulated in Sections 3.2.3 and 3.2.4. Because the focus of this study is on star scientists in general, the first consists of a sample of academic immunologists and the second of academic economists. By testing the hypotheses on academics in both the life and social sciences, I am able to extend the generalizability of the findings. I make use of the California salary disclosure policy, which makes public the salaries of all faculty employed in any state university. For this study, I solely focus on campuses in the University of California system that have an active immunology or economics department in 2008.

Immunologists

The University of California system consists of 10 campuses geographically spread across the state. Of these 10 campuses, five have immunology departments: Berkeley, Davis, Los Angeles, San Diego, and San Francisco. Manually collecting faculty rosters from each campus results in a list of 70 distinct immunologists. I clean this sample in two steps. First, using the *Frequently Occurring First Names and Surnames from the 1990 Census*,[6] I remove all immunologists whose surnames occur more than 0.05 % of the time in the population to ensure proper publishing count attribution. More common surnames are more likely to refer to two (or more) distinct individuals and as such erroneously inflate the immunologists human capital count. Removing the most common surnames reduces the sample from 70 immunologists to 65. Second, I remove all immunologists with

---

[6] http://www.census.gov/genealogy/names/

more than 50 years of publishing experience to ensure that senior or emeritus professors whose salary structures change, do not add noise to the (already) small sample. Three immunologists have more than 50 years of publishing experience. Removing them reduces the sample to 62. These 62 immunologists constitute the final immunologist sample.[7]

Economists

Within the 10 campuses of the University of California system, nine campuses contain stand-alone economics departments.[8] Culling department faculty rosters results in a list of 252 unique economists. Applying the same cleaning heuristic as with the immunologist sample, I remove all economists with surnames that occur more than 0.05% of the time in the United States population. Thirty-one economists have common surnames. Removing these economists reduces the sample to 221 economists. Similar to the immunologist sample, I remove economists with more than 40 years of publishing experience. The experience threshold is lower for economists as publishing in graduate school is less common. As the earliest age by which an economist typically graduates is 25,[9] 40 years of publishing experience corresponds approximately to a retirement age of 65. Eight economists have more than 40 years of publishing experience, which further reduces the sample to 213 economists.

3.3.2 Salary Data

The dependent variable of this study is the 2007 salaries of immunologists and economists in the University of California system. I obtain publicly available salary data from the

---

[7] Once I apply these two steps, the correlation between the rank of the commonness of the immunologist's surname and his publication count drops from -0.168 (a rank of 1 indicates the most common surname [Smith]) to -0.0532 (p-value of 0.6811).

[8] The San Francisco campus is a dedicated life and health science research university and as such does not have a conventional economics department.

[9] Most PhD graduates attend a four-year undergraduate institution where the average age at graduation is 21. Completing a PhD program in four years is fairly standard as well, resulting in an age of 25 at graduation.

Sacramento Bee newspaper.[10] The Sacramento Bee reports both an individual's *regular* pay in addition to *other* pay (summer money, etc.). I use the sum of these two values (*total* pay) for all analyses.[11]

3.3.3 Human Capital Data

I use productivity and helpfulness data to classify scientist as All-Stars, Lone Wolves, Mavens, or Non-Stars. Following prior work (Chapter 2), I utilize two different productivity measures: Impact Factor-weighted publications and citations. Both of these variables are available from Thomson ISI's *Web of Science* database. I collect Impact Factor-weighted publications by first retrieving all articles written by the sample of scientists in their respective set of journals. For immunologists, this involves the 140 journals ISI classifies as "immunology" journals, and for economists, the 209 "economics" classified journals. Using annual Impact Factor weights from ISI's *Journal Citations Report* service since 2000, I create average Impact Factor scores for each journal.[12] I then weight each scientist's publication by its journal's Impact Factor. The variable $IFPubs$ is the sum of Impact Factor-weighted publications published by the focal scientist by 2008. I also collect citations from ISI's *Web of Science*. The variable $Citations$ is the sum of all citations received by 2008[13] by all articles written by the focal scientist up until 2008.

I measure helpfulness as the number of academic acknowledgements a scientist receives. I collect immunology acknowledgements from the *Journal of Immunology*, the pre-eminent immunology journal within the field. Between 1998 and 2004 (inclusive), the *Journal of Immunology* published 17,233 articles. Using software and name identification algorithms (Councill, Giles, Han, and Manavoglu, 2005), I extract 60,481 acknowledge-

---

[10] http://www.sacbee.com/statepay/

[11] I also conduct the empirical analysis on only a scientist's base salary. The results are qualitatively similar, albeit with smaller point estimates but also smaller standard errors. As a result, statistical significance of the parameters of interest is largely unchanged.

[12] ISI started calculating journal Impact Factors in 2000.

[13] I count citations from all journal sources, not just economics or immunology journals.

ments from the 17,233 articles. From these 17,233, 264 acknowledgements go to the 62 immunologists in the sample.

I use the 1,343 articles that are available in machine-readable format from the *American Economic Review* between 1997 and 2005 (inclusive) to measure an economist's helpfulness.[14] I choose the *American Economic Review* because it is widely regarded as one of the top publishing outlets in the field, publishes both theoretical and empirical work, and does not suffer any institutional biases that other journals might.[15] From the 1,343 articles published, I extract 7,099 acknowledgements. Of these 7,099 acknowledgements, 212 are acknowledgements of economists in the sample.

Independent Variables

The main independent variables in this study are the four star types: All-Stars, Lone Wolves, Mavens, and Non-Stars. All-Stars are scientists with both high productivity and helpfulness; Lone Wolves are scientists with high productivity and average helpfulness; Mavens are scientists with average productivity and high helpfulness; Non-Stars are scientists with both average productivity and helpfulness. A natural decision arises from this taxonomy, namely, what cut-offs should be used to distinguish high and average scientists. In line with the prior literature, I use cut-offs of both the 90th and 95th percentiles. In addition, this study makes use of two productivity measures: Impact Factor-weighted publications and citations. Consequently, a scientist can be an Impact Factor-weighted publication-based star and/or a citation-based star. As a result, I can calculate the four star types – All-Stars, Lone Wolves, Mavens, and Non-Stars – four different ways:

1. Scientists above the 90th percentile of the acknowledgements and Impact Factor-

---

[14] The Papers and Proceedings of the American Economic Association meetings are published every year in the May issue of the *American Economic Review*. I include these articles in the analysis.

[15] The other often regarded top journals in economics are the *Quarterly Journal of Economics*, which is edited by Harvard University's economics department, and the *Journal of Political Economy*, which is edited by the University of Chicago's economics department.

weighted publications distributions have high productivity and helpfulness while those below are average.

2. Scientists above the 90th percentile of the acknowledgements and citations distributions have high productivity and helpfulness while those below are average.
3. Scientists above the 95th percentile of the acknowledgements and Impact Factor-weighted publications distributions have high productivity and helpfulness while those below are average.
4. Scientists above the 95th percentile of the acknowledgements and citations distributions have high productivity and helpfulness while those below are average.

The second main independent variable is the variable Internal. Internal is a measure of the share of a scientist's papers that has only one institution on the paper, indicating that all authors of the article are at the same institution. Internal $\in [0,1]$, where a value of 0 indicates that all papers written by the scientist have had at least one author at another institution and where a value of 1 indicates that all papers written by the scientist have been with coauthors at the same institution as the scientist himself. The measure acts as a proxy for the level of internal collaboration.

3.3.4 Descriptive Statistics

Figures 3.1 and 3.2 present Lorenz curves, which graphically depict the distribution of immunologist and economist salaries, Impact Factor-weighted publications, citations, and acknowledgements. As can be seen from both figures, all four distributions are highly skewed. The top 10% of immunologists capture over 20% of all salary, produce over 35% of Impact Factor-weighted publications, receive over 45% of citations, and receive over 50% of all acknowledgements. The Economist Lorenz curves are equally skewed. The top 10% of economists receive 22% of all salary, produce 33% of all Impact

Factor-weighted publications, receive 51% of citations, and receive just under 55% of all acknowledgements.

Figures 3.3 and 3.4 present scatter plots of the main productivity and helpfulness measures of both the immunologist and economist samples. I take logs of all the variables to remove the skewed nature of these data.[16] Across both samples, strong correlations exist between a scientist's Impact Factor-weighted publications (IF Pubs) and his citations. In addition, while salary appears to be positively correlated with Impact Factor-weighted publications and citations, the correlation is less strong for acknowledgements, albeit still positive.

Tables 3.1 and 3.2 present variable descriptive statistics for the immunologist and economist samples, respectively. As has been shown graphically in Figures 3.1 and 3.2, the data are heavily right skewed: all variable means are larger than their medians (except for Internal). The average salary of all immunologists in the sample is \$177,335.40, while the average immunologist has published just over 31 papers and has just over 21 years of experience. In addition, the average immunologist has written 46% of his papers where all coauthors are at the same institution. The average salary of economists is slightly lower at \$144,728.10. The average economist has written just over 16 papers in his career, has 18 years of publishing experience, and writes 56% of his papers with internal coauthors. Tables 3.3 and 3.4 present correlations between the variables described in Tables 3.1 and 3.2. Due to the high correlation between Impact Factor-weighted publications and citations, I do not include both measures in the same specification.

Table 3.5 presents by campus descriptive statistics of average immunologist salaries, publication counts, Impact Factor-weighted publication counts, citations, acknowledgements (Acks), and the percentage of papers written only with coauthors of the same institution as the scientist (Internal). The immunology sample reveals that while San Fran-

---

[16] To avoid dropping observations where acknowledgements are 0, I add 1 to the variable before I take the log.

cisco has the highest average citation and Impact Factor-weighted publications counts, immunologists at Los Angeles are paid the highest on average. Furthermore, behind San Francisco, Berkeley immunologists receive the most acknowledgements on average. With respect to internal collaboration, immunologists at all campuses write approximately 50% of their papers solely with internal colleagues, except for San Diego, where its immunologists write approximately 40% of their papers internally.

Table 3.6 presents characteristics of the economist sample by campus. The 33 economists in the economics department at Los Angeles receive the highest salary on average, yet economists at Berkeley receive the most citations, Impact Factor-weighted publications, and acknowledgements on average. Furthermore, economists at the Irvine and Santa Barbara campuses write approximately 70% of their papers with internal coauthors, while in contrast economists at Merced write approximately 30% of their papers with internal coauthors. Lastly, economists at both Irvine and Riverside publish articles in journals with an average Impact Factor of less than 1.[17]

As can be seen across both the immunologist and economist samples, clear variation across campuses exists for all covariates. To deal with this cross-campus variation empirically, I include campus fixed effects in all regressions. Consequently, I only use within campus variation to identify the parameters of interest. The next section describes the empirical strategy.

3.3.5 Methods

I empirically model the salary of a scientist to be a linear function of his star type, his experience, and the level of his internal collaboration. Formally:

---

[17] In 2007, two journals had Impact Factors of exactly 1, the *Journal of Agricultural Economics* and the *Review of International Political Economy.* These two journals were ranked 56th and 57th, respectively, by Impact Factor.

$$
\begin{aligned}
Y_{ic} &= \alpha + \beta_1 \text{All-Star}_{ic} + \beta_2 \text{Lone Wolf}_{ic} + \beta_3 \text{Maven}_{ic} + \beta_4 \text{Internal}_{ic} \\
&+ \beta_5 \text{Internal}_{ic} \cdot \text{All-Star}_{ic} + \beta_6 \text{Internal}_{ic} \cdot \text{Lone Wolf}_{ic} + \beta_7 \text{Internal}_{ic} \cdot \text{Maven}_{ic} \\
&+ X_{ic} + \phi_c + \varepsilon_{ic} \qquad (3.1)
\end{aligned}
$$

where Y_{ic} is the log salary of scientist i on campus c in 2007. X_{ic} is a vector of control variables, which in practice consists of the number of years of publishing experience for scientist i and its quadratic. I include campus fixed effects, represented by ϕ_c, to control for the large inter-campus heterogeneity observed in Tables 3.5 and 3.6. I use Ordinary least squares to estimate the linear relationship in Equation 3.1. I report Huber-White robust standard errors for all regression results.

β_1 through β_3 captures the relative salary difference between the three main star types and Non-Stars (the omitted category). Support for Hypothesis 1 would be found if $\beta_1 > 0; \beta_1 > \beta_3; \beta_2 > 0; \beta_2 > \beta_3$. Support for Hypothesis 2 would be found if $\beta_3 > 0$. Hypothesis 3 predicts a positive interaction between internal and Mavens. As such, Hypothesis 3 would receive support if $\beta_7 > 0$.

3.4 Results

I report three sets of results in this section. First, I present regression results of the specification outlined in Equation 3.1 both for the immunology sample (Tables 3.7 and 3.8) and the economist sample (Tables 3.9 and 3.10). I present robustness checks of the immunology sample where I define productivity by Impact Factor-weighted publications in Table 3.11 and results for when I measure productivity using citations in Table 3.12. I produce the same robustness checks for the economist sample and present them in Tables 3.13 and 3.14.

3.4.1 Main Regression Results

Immunologist Sample

Table 3.7 presents regression results predicting an immunologist's salary. The percentile cut-off for defining a star in this table is the 90th percentile.[18] Columns 1 through 3 define stars using Impact Factor-weighted publications as productivity measures, while Columns 4 through 6 measure productivity using citations.

The coefficients on All-Star and Lone Wolf are both statistically significant at the 5% level in Column 1. Because the All-Star, Lone Wolf, and Maven variables are dummy variables and because the dependent variable is a log, subtracting 1 from the antilog of the coefficient and multiplying by 100%[19] converts the coefficients into elasticities. As such, a Lone Wolf receives on average 106% more salary than a Non-Star (the omitted category), while an All-Star receives 33% more than a Non-Star. In addition, the salaries of All-Stars and Lone Wolves are both quantitatively larger than the salary of Mavens, but only the difference between salaries of Lone Wolves and Mavens is statistically distinct (at the 8% level). As a result, there is at least partial support for Hypothesis 1. The average salary of a Maven, however, is not statistically distinct from the salary of an average Non-Star, providing no support for Hypothesis 1. The size of the salary coefficients warrants some discussion as well. While the Lone Wolf coefficient appears much larger than the All-Star coefficient, the two coefficients are not statistically distinct from each other.

Column 2 introduces the internal variable, but the coefficient on Internal is statistically insignificant. The coefficients of All-Star and Lone Wolf, however, stay largely unchanged when controlling for Internal collaboration and retain their previous level of significance (5%). Column 3 introduces the interactions between Internal and the three

---

[18] That is, an All-Star is in the 90th percentile of both productivity and helpfulness, a Lone Wolf has productivity in the 90th percentile but helpfulness below the 90th percentile and a Maven has productivity below the 90th percentile but helpfulness in the 90th percentile. A Non-Star has both productivity and helpfulness below the 90th percentile.

[19] $(e^{\beta} - 1) \times 100\%$

star types, which allows for the testing of Hypothesis 3. Both All-Star and Lone Wolf retain their positive and statistically significant coefficients. However, because of the interactions, the level coefficients of the star types are much larger in magnitude. An All-Star that does not collaborate at all internally receives a salary 106% higher than a Non-Star, while a Lone Wolf receives a salary 542% higher than a a Non-Star. A Maven, on the other hand, who does not collaborate with anyone internally receives a salary 32% *less* than a Non-Star. However, a one standard deviation increase in the internal collaboration rate (0.22[20]) increases the salary of all stars by 11%, while a one standard deviation increase in Internal is associated with a 40%[21] increase in a Maven's salary relative to the salary of a Non-Star, providing support for Hypothesis 3.

Interestingly, Lone Wolves are compensated less when they collaborate more internally (as a percentage of all collaborations), providing speculative evidence that the Lone Wolves who are compensated the most collaborate most with the outside academic community. One explanation for this may simply be that because Lone Wolves are the most productive, they "match" with the best possible coauthors who are most likely not at their own institution. Conversely, Lone Wolves are also paid the highest, which is why we observe the negative correlation between pay and internal collaboration. The case of Mavens, however, is more interesting. Mavens do not have high productivity, and so the gains from spillovers that are generated internally through coauthorship may outweigh the loss in more optimal matching that would occur if the Maven collaborated externally. If this is the case, then campuses may be willing to pay Mavens (in the form of salary) to forego outside coauthorships (which would increase the Maven's direct productivity) in favor of internal coauthorships (which would increase the campuses' total output from the productivity spillovers generated by the Maven).

In Columns 4, 5, and 6, I classify a star using his citations as a proxy for his pro-

---

[20] From Table 3.1.

[21] $0.512 \cdot 0.22 + (e^{1.147 \cdot 0.22} - 1) = 40\%$

ductivity (instead of his Impact Factor-weighted publications). Column 4 provides very similar results to those shown in Column 1, whereby both All-Stars and Lone Wolves receive higher salaries than Non-Stars. The coefficient for Mavens is still statistically indistinguishable from 0 (Non-Stars). Column 6 introduces the interactions between star type and internal collaboration, and again we see a very similar pattern to when I define a star using Impact Factor-weighted publications. An increase of a scientist's internal collaboration is positively associated with salary, and an increase in internal collaboration is positively associated with salary for Mavens but negatively associated with salary for Lone Wolves.

Table 3.8 calculates star types using the 95th percentile as a cut-off instead of the 90th percentile which is used in Table 3.7. Column 1 indicates that the salaries of All-Stars are 34% higher than Non-Stars, the salaries of Lone Wolves are 223% higher than Non-Stars, and the salaries of Mavens are 45% higher than Non-Stars. The statistical significance of Mavens at the 95th percentile compared to their insignificance at the 90th percentile appears to indicate that only the far right tail of the helpfulness distribution is positively associated with salary. Including the effect of internal collaboration on salary increases the salaries for All-Stars and Lone Wolves and decreases the salary of Mavens. The impact of internal collaboration is also positively associated with salary, whereby a one standard deviation increase in internal collaboration is associated with an 11% increase in salary, *ceteris paribus*. Including the interactions between star types and internal collaboration is problematic, however, as only one Lone Wolf and one Maven exists at the 95th percentile, disallowing the inclusion of the interaction in the regression due to perfect collinearity. Columns 4 through 6 classify star types using citations as their measure of productivity. The results are very similar to those shown in Columns 1 through 3. Column 5 shows that internal collaboration is again positively associated with larger salaries. Meanwhile, All-Stars on average receive salaries that are 63% higher than those of Non-Stars. Lone Wolves receive 188% more salary than Non-Stars, while

Mavens receive salaries that are 34% higher than Non-Stars. The salary differential between Lone Wolves and Mavens is statistically significant at the 5% level, but the salary of Mavens is not statistically distinct from those of All-Stars. Unfortunately, only one immunologist is classified as an All-Star when I define high productivity as the 95th percentile of citations. As such, I cannot estimate an interaction effect for All-Stars. Lone Wolves, however, receive lower salaries when they collaborate more internally, and internal collaboration has no statistically significant impact on the salaries of Mavens.

As a whole, the immunologist sample provides medium to strong support for all predicted hypotheses. For all specifications, the salaries of All-Stars and Lone Wolves are higher than the salaries of Non-Stars. In addition, for all specifications (except Column 2 in Table 3.7), the salary of Lone Wolves is statistically larger than the salary of Mavens (Hypothesis 1). While Mavens do not receive statistically larger salaries than Non-Stars when I define stars using the 90th percentile cut-off, they do when classifying Mavens as immunologists in the 95th percentile of the helpfulness distribution (Hypothesis 2). Furthermore, the positive relationship between internal collaboration and salary is strong in most specifications, and I show support for the positive relationship between internal collaboration and salary for Mavens (Hypothesis 3) in Table 3.7.

Economist Sample

Tables 3.9 and 3.10 replicate the results from Tables 3.7 and 3.8 for the economist sample. Column 1 in Table 3.9 presents the base results between the association of star type and salary where I define star types by a 90th percentile cut-off and I measure productivity using Impact Factor-weighted publications. All-Stars receive salaries that are 57% higher than those of Non-Stars, Lone Wolves receive salaries that are 49% higher than Non-Stars, and Mavens receive salaries that are 30% higher than the salaries of Non-Stars. Including the level of internal collaboration in Column 2 for each economist has no discernible effect on the salary premiums for All-Stars, Lone Wolves, and Mavens reported in Column 1. Column 3 presents the interactions between internal collaboration and star type. No particular star type receives a different salary when they collaborate more internally than Non-Stars. Columns 4 through 6 provides results for star types where I measure productivity using citations. Column 4 presents similar magnitudes to those reported in Column 1, indicating that citations and Impact Factor publications make a minimal difference in predicting salary. Internal collaboration is not statistically associated with salary and does not change the coefficients of the star types in any meaningful way. Column 6 introduces the interactions between internal collaboration and star type. Neither All-Stars nor Mavens receive higher salaries than Non-Stars when they collaborate more internally, providing no support for Hypothesis 3, but Lone Wolves do receive lower salaries the more they collaborate internally, a finding previously discovered in the immunologist sample as well.

Table 3.10 defines stars using the 95th percentile cut-off of the productivity and helpfulness distributions. Column 1 indicates that All-Stars receive salaries 69% higher than Non-Stars. Lone Wolves, on the other hand, receive salaries that are 71% higher than Non-Stars. Mavens conversely receive salaries that are on average 24% higher than those of Non-Stars. Column 2 introduces the internal collaboration measure. Again, internal collaboration appears to have no impact on the salaries of economists. Column 3 introduces the interactions between internal collaboration and star type. Of all the star types, only All-Stars receive higher salaries when they collaborate more internally. A one standard deviation increase in internal collaboration (0.28[22]) is associated with a 172% increase in an All-Star's salary. Column 4 reveals similar results to those in Column 1, but the relative salary difference between All-Stars and Non-Stars is smaller when I define stars using citations. For Mavens, however, the salary difference is higher when I define stars using citations instead of Impact Factor-weighted publications. Column 6 introduces the interactions between internal collaboration and star types. I define

---

[22] Table 3.2.

only one economist as an All-Star when the productivity measure is citations and the star threshold is the 95th percentile. As such, I cannot estimate a coefficient on the interaction due to there not being any additional variation once I include the star type and internal collaboration variables. Nonetheless, Lone Wolves who collaborate more internally receive less salary.

Taken as a whole, the economist sample provides additional support, albeit slightly weaker than that provided by the immunologist sample, for the relationship between star type and salary.

3.4.2 Robustness Checks

This section provides robustness checks of the relationship between productivity, helpfulness, and salary. Tables 3.11 and 3.12 present regression evidence of the relationship between helpfulness and salary and between productivity and salary. Column 1 of Table 3.11 shows the relationship between acknowledgements and salary for the sample of immunologists. A one standard deviation increase in acknowledgements is associated with a 7.4% increase in salary. The log of Impact Factor-weighted publications are not statistically related to salary (Column 2) nor are they when I include acknowledgements (Column 3). In general, when examining the log-linear relationship between helpfulness and salary and the log-log relationship between Impact Factor-weighted publications and salary, little is significant, indicating that the relationship is not linear in nature. Column 6 shows the positive relationship between internal collaboration and salary. While the coefficients on the interactions between productivity and helpfulness are signed as expected – salary increases when higher productivity immunologists collaborate more externally and salary increases when more helpful immunologists collaborate more internally – neither coefficients are statistically significant.

Table 3.12 presents the same specifications as shown in Table 3.11 but uses the log of citations instead of the log of Impact Factor-weighted publications. Because of the strong correlation between IF pubs and citations, I cannot include both variables in the same specification. Again, little linear relationship exists between productivity (in this case citations) and salary. Internal collaboration, however, is again positively associated with salary.

Table 3.13 presents similar results to those of Table 3.11 but instead uses the economist sample. Column 1 shows that a one standard deviation increase in the acknowledgements of economists is associated with a 10.6% increase in salary. Column 2 shows that a 10% increase in Impact Factor-weighted publications corresponds to a 2.53% increase in salary. However, after controlling for productivity (IF pubs), the positive relationship between acknowledgements and salary disappears (Column 3). Including the level of internal collaboration increases the magnitude of the positive relationship between IFpubs and salary. In addition, an increase in internal collaboration is associated with an increase in salary. The interactions between internal collaboration and helpfulness (acknowledgements) and productivity (IFpubs) are not statistically distinct from 0.

Table 3.14 replicates Table 3.13 by including data on economist citations. A 10% increase in the number of citations an economist has received is associated with a 1.61% higher salary. The positive relationship between acknowledgements and salary holds, even after controlling for an economist's citations. Unfortunately, the relationship between internal collaboration and higher salaries disappears when I measure productivity using citations.

In general, these results reveal the appropriateness of examining the tails of the productivity and helpfulness distributions (stars), as simply relying on the linear relationship obfuscates the non-linearity in helpfulness, productivity, and their relationship with salary.

3.5 Discussion and Conclusion

This study attempts to determine the relative differences in wages for productivity and helpfulness star scientists. Due to the inherent inabilities of scientists to fully appropriate the benefits of their helpfulness, I theoretically predict that more productive scientists receive higher salaries than more helpful scientists. I find empirical support for these predictions in a sample of both academic immunologists and economists. Furthermore, the more internal collaboration a given scientist engages in, the higher his salary. The relationship between internal collaboration and salary is positive for helpfulness stars (Mavens) and negative for productivity stars (Lone Wolves). That is, the more Mavens collaborate with internal colleagues, the higher their salary, and the more Lone Wolves collaborate with external scientists, the higher their salary. This finding provides preliminary support for the notion that organizations will compensate Mavens for collaborating internally versus externally and thus generating positive spillovers for the organization. It also indicates that Lone Wolves are highly productive because they engage in the "best" collaborations possible, many of which fall outside the organization's boundaries. Organizations appear to not price spillovers that occur outside the organization's boundaries.

While strong evidence in favor of the hypothesis that observability of human capital drives a scientist's compensation, I am unable to directly test observability and instead rely on proxies. Future work calls for better measures of observable and unobservable characteristics of a scientist's human capital. Regardless, this study provides strong evidence of differences in compensation for helpfulness, a human capital characteristic that has been shown to have large and significant impact on the performance of others.

A tacit assumption of this study is that the benefits to organizations of productivity and helpfulness are homogenous. As shown in Figures 3.1 and 3.2, the distribution of scientists' shares of salary, acknowledgements, citations, and Impact Factor-weighted publications are all highly skewed. One explanation is that certain stars really are better suited for certain environments and are "worth" the higher pay. It also may be that the value to organizations is also highly skewed, whereby, for example, an idea that receives 10 times as many citations as another idea has an economic value in excess of 10 times. If this were the case, then there may not be overpayment but rather full appropriation on the part of the scientist (Rosen, 1981). This study attempts to account for this inherent skewed distribution in the quality of academic work by measuring a scientist's productivity using both citations and Impact Factor-weighted publications, but future work on determining the true value of academic output remains.

In addition, immunology and economics may be characterized as disciplines whereby discovering something first is of the utmost importance, somewhat akin to a patent race. If this were the case, then disproportionate rewards would be garnered by the institution to discover something first, and so despite diminishing returns, it still may make sense to pay for the "best" scientists (Tirole, 1993).

A natural follow-on question to the observation that Mavens capture less of their marginal productivity than Lone Wolves and All-Stars is why would someone still be helpful? While it is beyond the scope of this paper, work has been done in industrial psychology on Organizational Citizenry Behavior (OCB), indicating that helpfulness may indeed be a personality trait (Smith, Organ, and Near, 1983). Alternately, there also may be intrinsic rewards to participating in open science, of which being helpful plays a major role (Stern, 2004).

Chapter 2 identifies the important differences in the spillover impact of different star types. The empirical objective of this study is to determine whether or not these performance benefits for coauthors are spillovers, that is, unpriced benefits. By analyzing the salaries of All-Stars, Lone Wolves, and Mavens relative to Non-Stars, it appears that Lone Wolves are able to capture a large share of their productivity in the form of salary but that Mavens are often only payed slightly more than Non-Stars despite the positive performance impact they have on their coauthors (Chapter 2).

Peteraf (1993, pg. 187) famously wrote that "a Nobel Prize-winning scientist may be a unique resource, but unless he has firm-specific ties, his perfect mobility makes him an unlikely source of sustainable advantage." In two contexts, however, I demonstrated that due to the inability for Mavens to fully appropriate their marginal productivity – not because of firm-specific ties but due to differences in the observability of a scientist's full human capital – Mavens can act as important sources of organizational competitive advantage as they do indeed generate economically valuable human capital spillovers that are not fully appropriated by the scientist himself.

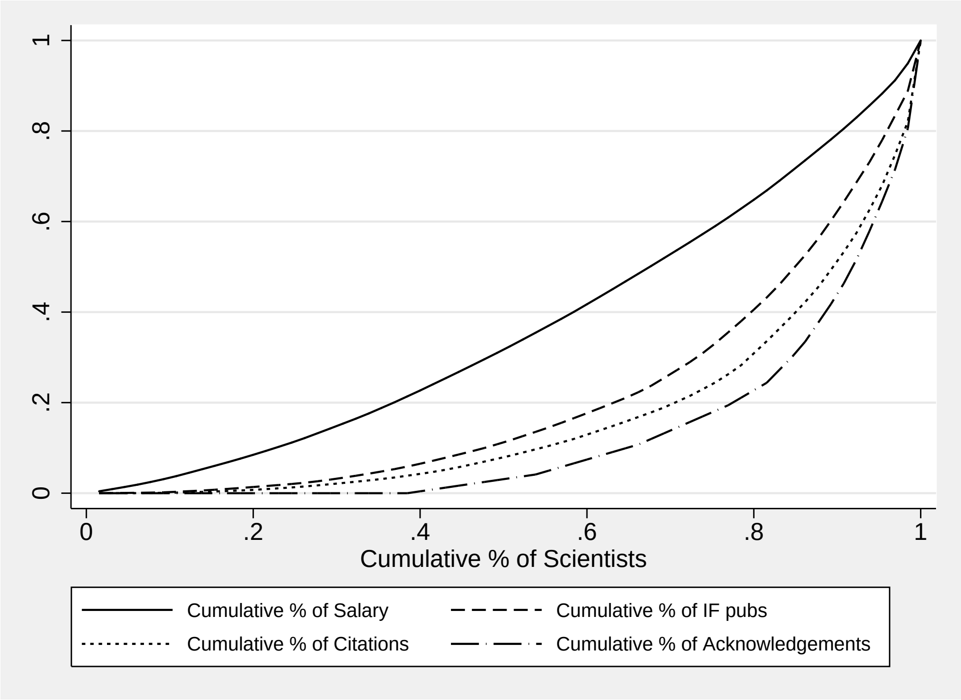

Figure 3.1: Immunologist Lorenz Curves

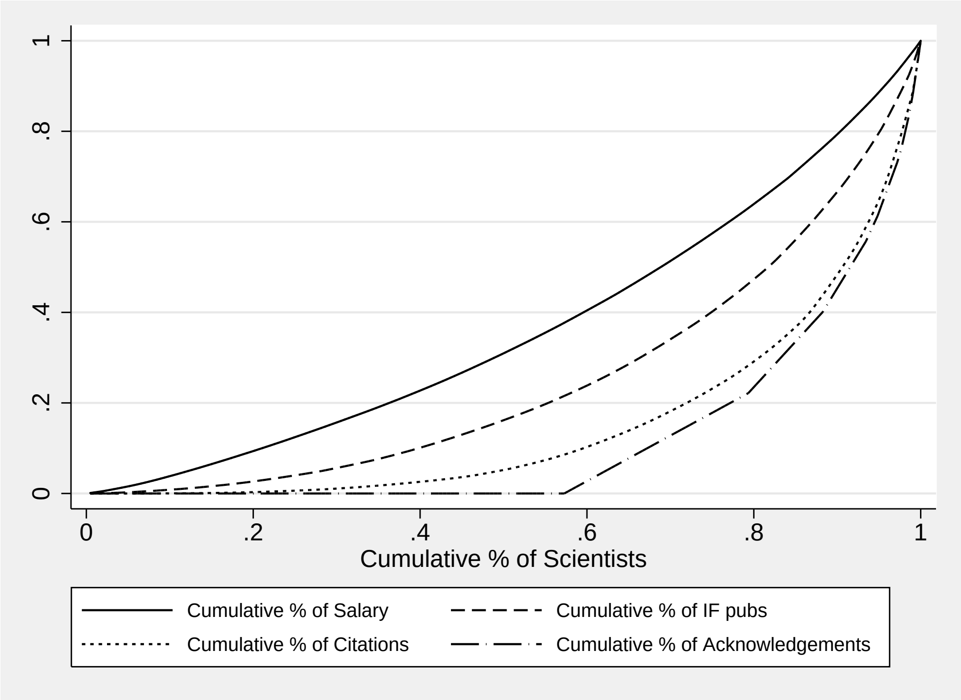

Figure 3.2: Economist Lorenz Curves

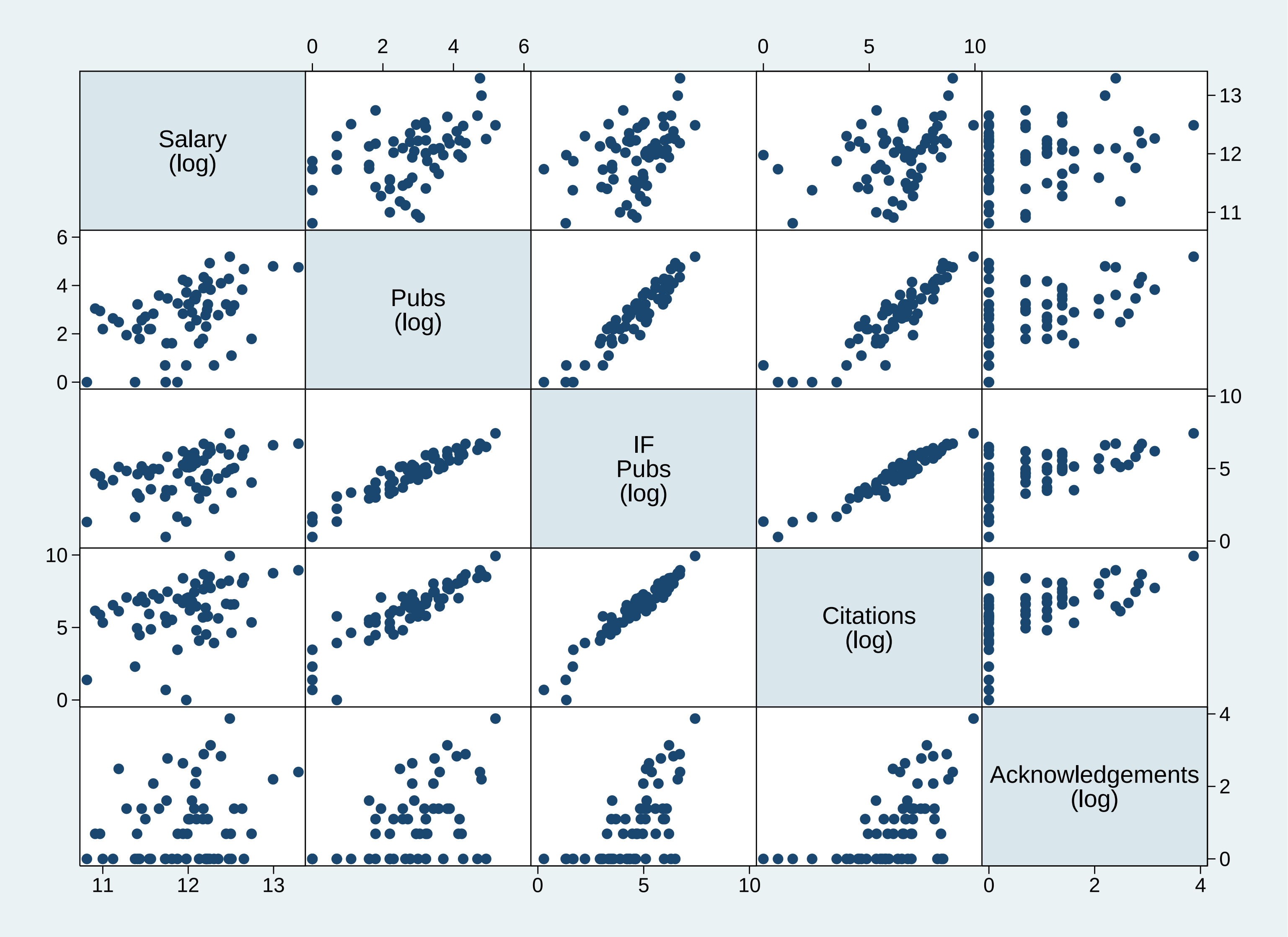

Figure 3.3: Immunologist Sample Scatter Plot Matrix: N = 62

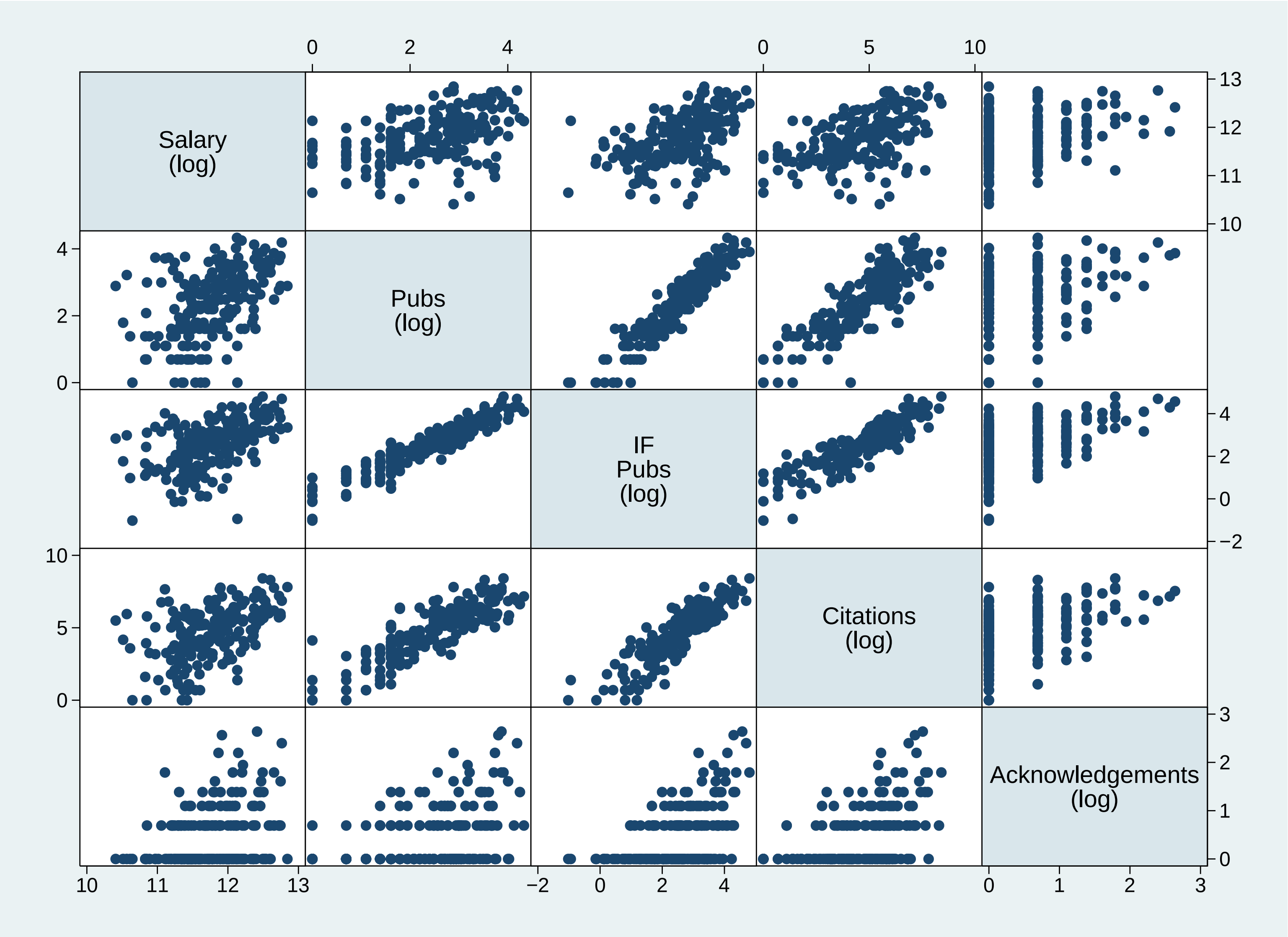

Figure 3.4: Economist Sample Scatter Plot Matrix: N = 213

Table 3.1: Immunology Sample Descriptive Statistics: N = 62

Variable	Mean	Std. Dev.	Min	p10	p25	p50	p75	p90	Max
Salary	177335.4	93922.9	49687.0	79099.0	108211.0	165299.5	209415.0	270256.0	591878.0
Publications	31.2	36.6	1.0	2.0	9.0	18.5	37.0	72.0	180.0
IF Publications	219.3	284.4	1.3	18.9	35.6	119.0	297.8	542.1	1688.7
Citations	1640.4	2985.6	1.0	60.0	211.0	728.5	1702.0	4429.0	20611.0
Acknowledgements	3.8	7.4	0.0	0.0	0.0	1.0	3.0	11.0	47.0
Experience	21.4	10.6	4.0	9.0	13.0	20.0	30.0	35.0	48.0
Internal	0.46	0.22	0.00	0.18	0.33	0.49	0.58	0.70	1.00

Table 3.2: Economics Sample Descriptive Statistics: N = 213

Variable	Mean	Std. Dev.	Min	p10	p25	p50	p75	p90	Max
Salary	144728.1	71551.9	33150.0	72887.0	92174.0	129203.0	182183.0	251041.0	378341.0
Publications	16.3	14.6	1.0	3.0	5.0	12.0	21.0	37.0	76.0
IF Publications	20.4	20.5	0.4	2.6	5.8	13.8	26.8	46.1	122.1
Citations	355.9	612.7	0.0	3.0	26.0	129.0	385.0	939.0	4500.0
Acknowledgements	1.0	1.9	0.0	0.0	0.0	0.0	1.0	3.0	13.0
Experience	18.0	11.4	0.0	3.0	8.0	17.0	27.0	34.0	40.0
Internal	0.56	0.28	0.00	0.14	0.38	0.57	0.75	0.95	1.00

Table 3.3: Immunology Sample Correlation Table: N = 62

	Variable	1	2	3	4	5	6
1	Salary (log)						
2	Publications (log)	0.44					
3	IF Publications (log)	0.41	0.94				
4	Citations (log)	0.38	0.90	0.96			
5	Acknowledgements	0.22	0.43	0.51	0.46		
6	Experience	0.38	0.65	0.57	0.54	0.18	
7	Internal	0.22	0.14	0.04	0.05	−0.04	0.33

Table 3.4: Economist Sample Correlation Table: N = 213

	Variable	1	2	3	4	5	6
1	Salary (log)						
2	Publications (log)	0.47					
3	IF Publications (log)	0.53	0.94				
4	Citations +1 (log)	0.50	0.87	0.89			
5	Acknowledgements	0.30	0.39	0.46	0.41		
6	Experience	0.25	0.65	0.57	0.69	0.19	
7	Internal	0.02	0.11	−0.00	0.12	0.01	0.45

Table 3.5: Immunologist Characteristics by School

	N	Salary	Pubs	IF Pubs	Citations	Acks	Internal
Berkeley	8	148864.2	28.5	314.3	1871.0	8.9	0.5
		(48232.2)	(22.9)	(248.9)	(1667.5)	(7.5)	(0.1)
Davis	14	129419.4	13.4	56.4	388.1	0.4	0.5
		(55383.6)	(12.0)	(49.9)	(407.0)	(0.6)	(0.3)
Los Angeles	8	234989.8	47.9	287.3	2071.2	4.0	0.5
		(160277.4)	(49.8)	(289.0)	(2734.1)	(4.1)	(0.2)
San Diego	25	194992.8	33.0	214.0	1525.9	1.7	0.4
		(94731.2)	(34.2)	(216.0)	(1781.5)	(3.5)	(0.2)
San Francisco	7	176752.7	43.9	377.8	3798.3	11.9	0.5
		(47849.5)	(61.5)	(592.5)	(7492.6)	(16.4)	(0.2)
Total	62	177335.4	31.2	219.3	1640.4	3.8	0.5
		(93922.9)	(36.6)	(284.4)	(2985.6)	(7.4)	(0.2)

Table 3.6: Economist Characteristics by School

	N	Salary	Pubs	IF Pubs	Citations	Acks	Internal
Berkeley	40	170780.4	23.1	37.6	837.7	1.9	0.6
		(80306.5)	(18.6)	(30.9)	(1092.6)	(2.8)	(0.3)
Davis	26	116756.5	13.8	15.2	182.8	1.0	0.6
		(39520.3)	(9.6)	(13.9)	(251.8)	(2.4)	(0.2)
Irvine	19	126480.2	20.7	18.9	252.2	0.6	0.7
		(51823.2)	(23.2)	(21.2)	(384.1)	(1.1)	(0.3)
Los Angeles	33	197760.6	14.0	18.5	248.2	0.8	0.5
		(84474.3)	(11.4)	(14.9)	(322.4)	(1.4)	(0.3)
Merced	3	126875.7	5.3	6.2	40.0	0.7	0.3
		(17761.7)	(5.8)	(6.6)	(52.2)	(1.2)	(0.3)
Riverside	15	121251.3	12.4	10.5	101.7	0.3	0.6
		(49814.2)	(11.7)	(11.8)	(119.8)	(0.5)	(0.2)
San Diego	36	127346.8	13.2	16.8	307.4	0.9	0.5
		(61362.3)	(11.8)	(14.9)	(394.9)	(1.7)	(0.3)
Santa Barbara	21	147514.2	15.7	17.9	367.9	0.8	0.7
		(76292.2)	(7.1)	(8.8)	(553.9)	(1.8)	(0.2)
Santa Cruz	20	107465.1	15.9	16.3	206.8	0.7	0.6
		(45818.6)	(15.4)	(15.3)	(251.9)	(0.9)	(0.3)
Total	213	144728.1	16.3	20.4	355.9	1.0	0.6
		(71551.9)	(14.6)	(20.5)	(612.7)	(1.9)	(0.3)

Table 3.7: Immunology Stars - 90th Percentile

Dependent Variable Productivity Variable	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) Citations	Salary (log) Citations	Salary (log) Citations
	(1)	(2)	(3)	(4)	(5)	(6)
All-Star	0.287*	0.328*	0.722+	0.334 * *	0.391 * *	0.836*
	(0.113)	(0.124)	(0.391)	(0.115)	(0.135)	(0.349)
Lone Wolf	0.723*	0.721*	1.860 * *	0.684 * *	0.684*	1.743 * *
	(0.301)	(0.344)	(0.266)	(0.242)	(0.272)	(0.239)
Maven	0.107	0.133	−0.390+	0.165	0.189	−0.332
	(0.174)	(0.155)	(0.229)	(0.144)	(0.130)	(0.248)
Experience	0.040	0.029	0.019	0.039	0.028	0.018
	(0.029)	(0.030)	(0.031)	(0.029)	(0.030)	(0.030)
Experience^2	−0.001	−0.000	−0.000	−0.001	−0.000	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Internal		0.329	0.512+		0.338	0.516+
		(0.260)	(0.270)		(0.262)	(0.273)
Internal · All-Star			−0.828			−0.891
			(0.820)			(0.748)
Internal · Lone Wolf			−2.075 * *			−1.996 * *
			(0.430)			(0.393)
Internal · Maven			1.147*			1.141*
			(0.508)			(0.559)
Intercept	11.276 * *	11.235 * *	11.239 * *	11.275 * *	11.232 * *	11.234 * *
	(0.314)	(0.303)	(0.309)	(0.314)	(0.302)	(0.308)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	62	62	62	62	62	62
Adjusted R^2	0.219	0.222	0.254	0.236	0.241	0.269
Linear Tests						
Maven = Lone Wolf	0.081	0.123	0.000	0.048	0.081	0.000
Maven = All-Star	0.344	0.270	0.001	0.254	0.170	0.000

Robust standard errors in parentheses.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

p-values of the null hypothesis that the two values are equal are reported for linear tests.

Table 3.8: Immunology Stars - 95th Percentile

Dependent Variable Productivity Variable	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) Citations	Salary (log) Citations	Salary (log) Citations
	(1)	(2)	(3)	(4)	(5)	(6)
All-Star	0.289 ∗ ∗	0.369 ∗ ∗	0.921 ∗ ∗	0.369 ∗ ∗	0.490 ∗ ∗	0.505 ∗ ∗
	(0.103)	(0.131)	(0.306)	(0.101)	(0.108)	(0.112)
Lone Wolf	1.173 ∗ ∗	1.401 ∗ ∗	1.406 ∗ ∗	0.997 ∗ ∗	1.059 ∗ ∗	1.654 ∗ ∗
	(0.180)	(0.217)	(0.220)	(0.163)	(0.237)	(0.330)
Maven	0.375 ∗ ∗	0.332 ∗ ∗	0.311∗	0.312∗	0.294∗	−0.233
	(0.111)	(0.117)	(0.125)	(0.134)	(0.130)	(0.401)
Experience	0.037	0.019	0.018	0.037	0.022	0.018
	(0.030)	(0.030)	(0.030)	(0.029)	(0.029)	(0.030)
Experience^2	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Internal		0.538∗	0.552∗		0.445+	0.511+
		(0.255)	(0.261)		(0.234)	(0.257)
Internal · All-Star			−1.174+			
			(0.658)			
Internal · Lone Wolf						−1.444∗
						(0.655)
Internal · Maven						0.830
						(0.570)
Intercept	11.265 ∗ ∗	11.223 ∗ ∗	11.238 ∗ ∗	11.292 ∗ ∗	11.258 ∗ ∗	11.264 ∗ ∗
	(0.316)	(0.301)	(0.307)	(0.315)	(0.304)	(0.310)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	62	62	62	62	62	62
Adjusted R^2	0.220	0.252	0.239	0.265	0.283	0.267
Linear Tests						
Maven = Lone Wolf	0.000	0.000	0.000	0.001	0.005	0.000
Maven = All-Star	0.307	0.769	0.127	0.703	0.207	0.042

Robust standard errors in parentheses.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

p-values of the null hypothesis that the two values are equal are reported for linear tests.

Table 3.9: Economist Stars - 90th Percentile

Dependent Variable Productivity Variable	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) Citations	Salary (log) Citations	Salary (log) Citations
	(1)	(2)	(3)	(4)	(5)	(6)
All-Star	0.453*	0.455*	1.422+	0.433*	0.434*	1.429+
	(0.188)	(0.185)	(0.743)	(0.188)	(0.185)	(0.736)
Lone Wolf	0.402 **	0.403 **	0.612 **	0.394 **	0.393 **	0.900 **
	(0.107)	(0.105)	(0.198)	(0.117)	(0.114)	(0.274)
Maven	0.266 **	0.254 **	0.263	0.272 **	0.260 **	0.333+
	(0.092)	(0.096)	(0.215)	(0.091)	(0.094)	(0.198)
Experience	0.032 **	0.033 **	0.031 **	0.031 **	0.033 **	0.029 **
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Experience^2	−0.001 **	−0.001 **	−0.001*	−0.001 **	−0.001 **	−0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Internal		−0.130	−0.090		−0.122	−0.053
		(0.121)	(0.129)		(0.118)	(0.124)
Internal · All-Star			−1.461			−1.518
			(1.131)			(1.132)
Internal · Lone Wolf			−0.314			−0.767+
			(0.337)			(0.410)
Internal · Maven			−0.004			−0.126
			(0.358)			(0.330)
Intercept	11.477 **	11.521 **	11.523 **	11.500 **	11.542 **	11.544 **
	(0.109)	(0.120)	(0.123)	(0.109)	(0.120)	(0.121)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	213	213	213	213	213	213
Adjusted R^2	0.241	0.242	0.241	0.243	0.242	0.248
Linear Tests						
Maven = Lone Wolf	0.273	0.232	0.177	0.354	0.312	0.062
Maven = All-Star	0.342	0.302	0.134	0.412	0.370	0.150

Robust standard errors in parentheses.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
p-values of the null hypothesis that the two values are equal are reported for linear tests.

Table 3.10: Economist Stars - 95th Percentile

Dependent Variable Productivity Variable	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) IF pubs	Salary (log) Citations	Salary (log) Citations	Salary (log) Citations
	(1)	(2)	(3)	(4)	(5)	(6)
All-Star	0.523 ∗∗	0.507 ∗∗	−1.368 ∗∗	0.457 ∗∗	0.451 ∗∗	0.418 ∗∗
	(0.179)	(0.184)	(0.495)	(0.104)	(0.102)	(0.105)
Lone Wolf	0.535 ∗∗	0.532 ∗∗	0.467 ∗∗	0.319	0.333+	1.878 ∗∗
	(0.123)	(0.122)	(0.165)	(0.202)	(0.195)	(0.371)
Maven	0.218∗	0.210∗	0.018	0.345 ∗∗	0.329∗	0.093
	(0.097)	(0.106)	(0.171)	(0.130)	(0.135)	(0.166)
Experience	0.034 ∗∗	0.035 ∗∗	0.035 ∗∗	0.034 ∗∗	0.036 ∗∗	0.034 ∗∗
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Experience^2	−0.001 ∗∗	−0.001 ∗∗	−0.001 ∗∗	−0.001 ∗∗	−0.001 ∗∗	−0.001 ∗∗
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Internal		−0.106	−0.122		−0.137	−0.082
		(0.121)	(0.128)		(0.122)	(0.124)
Internal · All-Star			3.574 ∗∗			
			(1.038)			
Internal · Lone Wolf			0.115			−2.105 ∗∗
			(0.339)			(0.599)
Internal · Maven			0.407			0.506
			(0.285)			(0.319)
Intercept	11.490 ∗∗	11.527 ∗∗	11.522 ∗∗	11.510 ∗∗	11.557 ∗∗	11.577 ∗∗
	(0.109)	(0.121)	(0.124)	(0.110)	(0.119)	(0.120)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	213	213	213	213	213	213
Adjusted R^2	0.234	0.233	0.226	0.214	0.214	0.239
Linear Tests						
Maven = Lone Wolf	0.045	0.049	0.024	0.914	0.986	0.000
Maven = All-Star	0.134	0.159	0.007	0.471	0.446	0.108

Robust standard errors in parentheses.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
p-values of the null hypothesis that the two values are equal are reported for linear tests.

Table 3.11: Robustness Checks: Immunology (Impact Factor-weighted publications)

Dependent Variable	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)
	(1)	(2)	(3)	(4)	(5)	(6)
Acks	0.010+ (0.006)		0.007 (0.006)	−0.077 (0.065)	0.009+ (0.005)	−0.018 (0.022)
Experience	0.044 (0.030)	0.035 (0.033)	0.036 (0.033)	0.035 (0.033)	0.013 (0.038)	0.004 (0.041)
Experience^2	−0.001 (0.001)	−0.000 (0.001)	−0.001 (0.001)	−0.001 (0.001)	−0.000 (0.001)	−0.000 (0.001)
IF Pubs (log)		0.063 (0.056)	0.043 (0.055)	0.049 (0.055)	0.073 (0.062)	0.175 (0.110)
Acks · IF Pubs				0.012 (0.009)		
Internal					0.496 (0.305)	1.020* (0.409)
Internal · IF Pubs						−0.172 (0.123)
Internal · Acks						0.061 (0.048)
Intercept	11.186 * * (0.320)	11.057 * * (0.338)	11.082 * * (0.340)	11.198 * * (0.342)	10.943 * * (0.315)	10.648 * * (0.355)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	62	62	62	62	62	62
Adjusted R^2	0.167	0.166	0.158	0.161	0.178	0.164

Robust standard errors in parentheses.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3.12: Robustness Checks: Immunology (Citations)

Dependent Variable	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)
	(1)	(2)	(3)	(4)	(5)	(6)
Acks	0.010+		0.008	−0.052	0.010+	−0.018
	(0.006)		(0.006)	(0.055)	(0.005)	(0.024)
Experience	0.044	0.035	0.037	0.037	0.016	0.003
	(0.030)	(0.033)	(0.033)	(0.033)	(0.038)	(0.041)
Experience^2	−0.001	−0.000	−0.001	−0.001	−0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Citations (log)		0.041	0.026	0.025	0.043	0.131+
		(0.041)	(0.041)	(0.041)	(0.044)	(0.077)
Acks · Citations				0.006		
				(0.006)		
Internal					0.466	1.164 ∗∗
					(0.296)	(0.434)
Internal · Citations						−0.154
						(0.093)
Internal · Acks						0.063
						(0.051)
Intercept	11.186 ∗∗	11.094 ∗∗	11.111 ∗∗	11.213 ∗∗	10.999 ∗∗	10.655 ∗∗
	(0.320)	(0.344)	(0.347)	(0.347)	(0.326)	(0.343)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	62	62	62	62	62	62
Adjusted R^2	0.167	0.163	0.156	0.154	0.173	0.166

Robust standard errors in parentheses.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3.13: Robustness Checks: Economics (Impact Factor-weighted publications)

Dependent Variable	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)
	(1)	(2)	(3)	(4)	(5)	(6)
Acks	0.056 ∗ ∗ (0.015)		0.019 (0.014)	−0.003 (0.065)	0.017 (0.014)	0.025 (0.045)
Experience	0.030 ∗ ∗ (0.010)	0.000 (0.010)	0.000 (0.010)	0.000 (0.010)	−0.006 (0.010)	−0.007 (0.010)
Experience^2	−0.001∗ (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
IF Pubs (log)		0.253 ∗ ∗ (0.041)	0.236 ∗ ∗ (0.045)	0.235 ∗ ∗ (0.045)	0.264 ∗ ∗ (0.044)	0.338 ∗ ∗ (0.081)
Acks · IF Pubs				0.006 (0.015)		
Internal					0.209+ (0.114)	0.457+ (0.239)
Internal · IF Pubs						−0.119 (0.119)
Internal · Acks						−0.016 (0.084)
Intercept	11.506 ∗ ∗ (0.107)	11.142 ∗ ∗ (0.115)	11.153 ∗ ∗ (0.117)	11.153 ∗ ∗ (0.117)	11.037 ∗ ∗ (0.129)	10.887 ∗ ∗ (0.183)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	213	213	213	213	213	213
Adjusted R^2	0.236	0.355	0.356	0.353	0.362	0.362

Robust standard errors in parentheses.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3.14: Robustness Checks: Economics (Citations)

Dependent Variable	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)	Salary (log)
	(1)	(2)	(3)	(4)	(5)	(6)
Acks	0.056 ∗ ∗ (0.015)		0.027+ (0.015)	0.004 (0.067)	0.026+ (0.014)	0.037 (0.048)
Experience	0.030 ∗ ∗ (0.010)	−0.019+ (0.011)	−0.018 (0.011)	−0.018 (0.011)	−0.023+ (0.012)	−0.024∗ (0.012)
Experience^2	−0.001∗ (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Citations (log)		0.161 ∗ ∗ (0.023)	0.149 ∗ ∗ (0.024)	0.148 ∗ ∗ (0.024)	0.158 ∗ ∗ (0.025)	0.186 ∗ ∗ (0.042)
Acks · Citations				0.003 (0.010)		
Internal					0.132 (0.117)	0.317 (0.207)
Internal · Citations						−0.049 (0.057)
Internal · Acks						−0.021 (0.089)
Intercept	11.506 ∗ ∗ (0.107)	11.247 ∗ ∗ (0.103)	11.249 ∗ ∗ (0.104)	11.249 ∗ ∗ (0.104)	11.186 ∗ ∗ (0.116)	11.098 ∗ ∗ (0.148)
Campus F.E.	✓	✓	✓	✓	✓	✓
Observations	213	213	213	213	213	213
Adjusted R^2	0.236	0.344	0.350	0.347	0.351	0.348

Robust standard errors in parentheses.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Chapter 4

Asymmetric Information and Labor Mobility: The Case of Scientist Productivity

4.1 Introduction

One of the prime mechanisms for knowledge transfer is through the mobility of individuals. Indeed, an active and growing literature that focuses on the relationship between the mobility of knowledge workers and knowledge spillovers finds strong support for this claim (Stephan, 1996; Almeida and Kogut, 1999; Rosenkopf and Almeida, 2003; Zellner, 2003; Agrawal, Cockburn, and McHale, 2006; Oettl and Agrawal, 2008). While the relationship between mobility and the diffusion of knowledge has been well documented empirically, scholars have placed less focus on the characteristics of the mobile. While a subset of the field has examined how the histories and characteristics of movers affect organizational level outcomes (Song, Almeida, and Wu, 2003), little attention has been given to examining what characteristics predict mobility. We know that productivity stars are important for generating ideas (Lotka, 1926) and that helpfulness stars are important for generating spillovers (Chapters 2 and 3), but are stars more or less mobile than non-stars? What is the relationship between an individual's productivity and his mobility?

To find out, I develop a simple model whereby the mobility decision of an individual hinges on the value ascribed to him by outside institutions. Outside institutions, however, are only able to value the individual using universally observable characteristics, while the individual's current institution is able to value the individual using his characteristics that are only observable within the institution in conjunction with the same universally available information the outside institutions have access to. As such, I predict that an individual's mobility increases with his observable productivity but decreases with his unobservable productivity. By separating observable from unobservable characteristics, I am able to not only provide insight into the mobility decisions of productivity stars but also helpfulness stars, whose productivity (spillovers to others) may be more difficult to observe externally.

Using publication and citation data from ISI's *Web of Science* to measure observable productivity as well as academic acknowledgement counts from the *Journal of Immunology* to measure unobservable productivity, I construct a panel dataset of 4,665 immunologists and track their mobility between 1974 and 1989 as a function of their observable and unobservable productivity. I find that a one standard deviation increase in an immunologist's unobservable productivity is associated with a 0.6% decrease in mobility on average, while a one standard deviation increase in an immunologist's observable productivity is associated with a 1% increase in mobility on average.

The next section provides a literature review, the model, and the hypotheses to be empirically tested. Section 4.3 introduces the data, describes the sample construction, presents descriptive statistics, and outlines the empirical strategy. Section 4.4 discusses the main results, checks for robustness, and provides some supplementary analysis. Finally, Section 4.5 provides discussion and concluding thoughts.

4.2 Literature and Theory

The study of the relationship between labor mobility and performance is well established within the strategic management literatures. Almeida and Kogut (1999) present one of the earliest empirical accounts of the relationship between labor mobility and the localization of knowledge spillovers. In more recent work, Oettl and Agrawal (2008) find that knowledge flow externalities are generated by the international movement of inventors. Song, Almeida, and Wu (2003) provide empirical evidence of firm learning from the hiring of engineers – the so-called “learning-by-hiring” effect. Relatedly, Rosenkopf and Almeida (2003) show that inventor mobility increases inter-firm knowledge flows above and beyond any gains that accrue to the firm from geographic propinquity. All of these papers, as with much of the labor mobility literature within strategy, treat the source or cause of labor mobility as exogenous and instead focus solely on labor mobility’s impact on performance rather than its causes in the first place. Furthermore, the majority of the literature relating labor mobility to firm level performance outcomes treats the human capital of the mobile as homogenous. While economists have developed a sizable theoretical literature on the causes of labor mobility and how mobility varies by human capital (Lazear, 1986; MacDonald, 1988; Topel, 1991; Topel and Ward, 1992), the majority of this body of work, however, is theoretical, thus welcoming and encouraging empirical inquiry.

Within the strategy literature, scholars have given little attention to the determinants of mobility. One notable exception is the work of Marx, Strumsky, and Fleming (2009), who examine the impact of noncompete agreements on the restriction of mobility. Using the state of Michigan’s 1985 reversal of its enforcement of its noncompete legislation, the authors find that indeed the presence of noncompete covenants restrict the movement of labor. In addition, this labor mobility attenuation is strongest for inventors with firm-specific skills and specialized expertise. Lastly, the authors find that inventors with more patents are less likely to move.

This negative relationship between an individual's productivity has been found in numerous other studies. Hoisl (2007) finds that an inventor's mobility is decreasing in his productivity even after controlling for the reverse causality of mobility positively influencing his productivity. Similar to the findings of Hoisl (2007), Schankerman, Shalem, and Trajtenberg (2006) find that more productive software inventors are less likely to move, but if they do move they are more likely to move to contexts, such as smaller firms and start-ups, in which they will be more able to capture their marginal productivity.

Other scholars have found a similar negative relationship between individual productivity and mobility in related literature on the effect of labor mobility on entrepreneurship. Groysberg, Nanda, and Prats (2007) study the mobility decisions of high performing (star) sell-side analysts. They find that, conditional on movement, stars are more likely to move to entrepreneurial ventures, in line with the findings of Schankerman, Shalem, and Trajtenberg (2006). Overall, though, star analysts are less likely to move in general, presumably due to the high context and firm specificity of their disproportionately high productivity. In a study on the impact of work experience in firms of varying sizes on the entrepreneurship decision, Elfenbein, Hamilton, and Zenger (2008) find a similar negative relationship between employee earnings and employee mobility.

While the body of literature examining the causes and consequences of labor mobility focuses on knowledge workers, it primarily surveys the actions of engineers and inventors, largely leaving scientists overlooked. Yet we know that scientific links are important for firm innovation (Cockburn and Henderson, 1998). Zucker, Darby, and Torero (2002) examine the mobility decisions of biotechnology stars and find that more productive stars move from academic institutions to for-profit firms at a faster rate than less productive stars. This latter paper is a welcome step in the direction of understanding the causes of scientist mobility, yet the authors' setting is still very specific (star moves from academia to private-sector firms), and they still only focus on observable facets of an individual's productivity.

The goal of this chapter is much simpler: What is the relationship between a scientist's productivity and his mobility in general? In the next section, I develop a simple analytical model that examines the mobility decision as a function of a scientist's productivity. Instead of examining how a scientist's mobility is moderated by the generalizability of his skill, I examine how the relationship between observable and unobservable productivity affects scientific mobility. I find that mobility is increasing in a scientist's observable productivity but decreasing in his unobservable productivity. In addition, the model shows that, depending on their relative levels, observable and unobservable productivity can be both complementary and substitutionary in nature. The theoretical work of Lazear (1986) is very similar in nature to my study's theoretical motivation. Lazear (1986) examines the extent to which heterogenous worker ability influences the outside offers a worker receives and subsequently the likelihood of being "raided" by an outside firm finding that on average workers who are "raided" have higher general productivity while workers who are unsuccessfully "raided" (that is, they receive outside offers but do not move) have higher firm-specific productivity. Lazear (1986) focuses on firm-specific skill, whereas this study focuses on two types of skill, both of which are generalizable,[1] but one is observable and one is not. This is an important distinction, as observability, rather than the underlying skill per-se, is the source of the information asymmetry and the (possibly) inefficient sorting of individuals who can act as a source of competitive advantage for organizations.

4.2.1 A Simple Model

A scientist will move to a new institution from his current institution when his value at the new institution is greater than his value at his current institution. That is, $V_n > V_c$, where the subscripts n and c refer to the scientist's new and current institutions, respectively. I assume that a scientist's total productivity solely determines his value.

---

[1] That is, this productivity has no firm-specific component.

Total productivity, $\rho \geq 0$, is made up of two components, direct productivity ($\delta \geq 0$) and indirect productivity ($\iota \geq 0$), such that $\rho = \delta + \iota$.

A scientist's direct productivity captures the aspects of his productivity, such as publications and citations. In general, direct productivity captures the content of the scientist's production that is easily observable (and priced) by the market and that has a direct impact on his own performance. Indirect productivity, on the other hand, captures aspects of the scientist's productivity that benefits the current institution in ways above and beyond the benefits (if any) generated by his direct productivity. Indirect productivity includes facets of the scientist's productivity that are difficult to observe outside of the current institution, such as teaching quality, service, helpfulness, and collegiality.[2]

The scientist's current institution can observe his total productivity, $\delta + \iota$, while the new institution can only observe his direct productivity, δ. If we assume that a scientist's value is proportional to his productivity, then the scientist's current institution will value the scientist as follows:

$$V_c = \delta + \iota$$

The value of a scientist for the new institution comes as a result of two assumptions. First, I assume that because of limited information and uncertainty over future performance, the matching process between institutions and scientists is suboptimal, and as such, the value of a scientist for new institutions will be larger than for the current institution. This assumption is in line with prior work on wages and mobility (Jovanovic, 1979; Jovanovic and Moffitt, 1990). Second, following the work of Rosen (1981), I assume

---

[2] Agrawal, Cockburn, and McHale (2006) argue that individuals will make investments in geographically localized social capital up to the point where the marginal cost of these investments equals the marginal benefit of these connections. As such, individuals will invest less in social capital if they expect to move frequently (as mobility decreases the value of localized social capital) . The concerns in the context of this study are that similar behavior exists with helpfulness and that the mobility of an individual determines an individual's helpfulness instead of the reverse. While I am unable to separately identify empirically what level of helpfulness is local versus global, the helpfulness measure described in Section 4.3 is not a within-organization construct. As such, I examine an individual's global helpfulness instead of only his helpfulness at his local institution. Doing so reduces this concern of reverse causality.

that a scientist's value for new institutions is a strictly increasing convex function of direct productivity. For simplicity, this convex function will take on the form of a simple quadratic. It follows that the value of a scientist for other institutions is:

$$V_n = \delta^2$$

The likelihood of a move occurring can be modeled as:

$$\Pr(Move = 1) = \frac{V_n}{V_n + V_c}$$

so that when the old and new institutions' values are the same, the probability of moving is 0.5. Substituting in the values for the new and current institutions results in the mobility equation:

$$M = \frac{\delta^2}{\delta^2 + \delta + \iota} \tag{4.1}$$

Taking the first derivative from equation 4.1 with respect to direct productivity reveals the change in mobility probability by the change in direct productivity:

$$\frac{\partial M}{\partial \delta} = \frac{\delta(\delta + 2\iota)}{(\delta + \delta^2 + \iota)^2} > 0 \tag{4.2}$$

The likelihood of a scientist's mobility is increasing in direct productivity (δ), thus generating our first hypothesis.

Hypothesis 1 (H1) The larger the direct productivity of an individual, the greater the likelihood of mobility

Looking at the change in mobility with a change in indirect productivity (ι) gives us our second testable hypothesis:

$$\frac{\partial M}{\partial \iota} = -\frac{\delta^2}{(\delta + \delta^2 + \iota)^2} < 0 \tag{4.3}$$

Hypothesis 2 (H2) The larger the indirect productivity of an individual, the lower the likelihood of mobility.

I now turn to the question of whether or not direct and indirect productivity are substitutes or complements. Thus, I take the cross partial derivative of M:

$$\frac{\partial^2 M}{\partial \iota \partial \delta} = \frac{2(\delta^3 - \delta\iota)}{(\delta + \delta^2 + \iota)^3} \gtrless 0 \tag{4.4}$$

The impact of an increase in ι and δ on the likelihood of mobility is ambiguous. The impact of the interaction depends on the relative size of ι to δ. When $\iota > \delta^2$, the likelihood of mobility increases. It conversely decreases when $\iota < \delta^2$. Intuitively, the increase in labor mobility that arises out of larger direct productivity is relative to the scientist's indirect productivity. Taken together and adopting the new star scientists taxonomy developed in Chapter 2, mobility increases for Lone Wolves, and decreases for Mavens.

4.3 Data and Empirical Strategy

To test the hypotheses developed in Section 4.2.1, I construct a sample of academic immunologists and examine the impact of direct and indirect productivity measures on the likelihood of movement. I use data from the ISI *Web of Science* to measure an immunologist's direct productivity and to discern mobility. I measure direct productivity using an immunologist's citations and counts of his Impact Factor-weighted publications. Building off the work developed in Chapter 2, I measure indirect productivity using a count of acknowledgements received in the *Journal of Immunology.*

The study calls for identifying the movements of immunologists across institutions. Many studies have made use of patent data from the United States Patent and Trademark Office to infer mobility (Agrawal, Cockburn, and McHale, 2006; Fleming, Mingo, and Chen, 2007; Oettl and Agrawal, 2008). The clear benefits of using patent data to infer mobility is that on each patent, location attributes are available for every inventor.

While these data would be appropriate for patenting scientists, very few immunologists patent. Conversely, we can cast our attention to immunology publications. The most common data source for science publications is the ISI *Web of Science*. Unfortunately, the *Web of Science* does not provide a one-to-one mapping of institutional affiliations (and by extension geography) to authors listed on an article.[3] That is, the *Web of Science* provides a list of all institutions represented per article as well as a list of all authors but provides no mechanism for matching a specific author to a specific institution. As such, without knowing the exact location of an author at any given time, it is impossible to infer change in institutional affiliation (movement) using these raw data. To overcome this, I develop an institution attribution algorithm, which I discuss in the next section.

4.3.1 Institution Attribution

I develop an algorithm for identifying an author's location using data from the *Web of Science* in the absence of a direct link identifying which institution(s) each author is affiliated with at a given point in time. While the *Web of Science* does not provide the aforementioned direct link, data are available that can help in ascribing institutions to authors. First, the data contain a list of authors in the order in which they appear on the original manuscript. Second, the data contain a list of institutions also in the order in which they appear on the original manuscript. Lastly, over half of all articles contain a reprint author, which for at least one author provides a direct mapping between author name and institution. Using these data, the algorithm works as follows. From the set of 116,305 articles written in immunology journals[4] between 1973[5] and 1990, I attribute a specific institution to a particular author if one of the following conditions is met:

1. If the immunologist is the reprint author of the article, then I assign the reprint

---

[3] The *Web of Science* has linked authors to specific institutions on papers published since 2008.

[4] As defined by the *Web of Science* list of immunology journals.

[5] Affiliation data from the ISI are unreliable prior to 1973

affiliation to the reprint author.[6]

2. If the immunologist is the first author and no reprint author is listed, then I assign the first affiliation to the first author.[7]

3. If the immunologist is on an article that has only one institution, then I affiliate *all* authors on the article to the single institution.

Th algorithm relies on frequent publishing to match along one of the three aforementioned criteria. The field of immunology is especially well suited to this as the annual publishing rates for immunologists are very high (on average over 2.7 publications[8]).

4.3.2 Movement

The institution attribution algorithm attempts to assign institutions to all immunologists who published between 1973 and 2007. Identifying movements between two institutions across time, however, requires additional logic. First, it may be that an immunologist appears to be affiliated with multiple institutions in the same year. This can happen for numerous reasons. Most commonly, this results from one-year sabbaticals.[9] Second, it may be that variations in the spelling of institutions generates the impression of a move, when in fact it is a false positive error. To deal with these "non-moves," I adopt the following heuristic. First, I remove all instances of moves where I observe that an individual immediately returns to a previous institution.[10] Second, I normalize the

---

[6] 66,389 of the 116,305 articles between this time period identify a reprint author.

[7] If a reprint author is listed, then the assumption of mapping the first listed author to the first listed affiliation no longer holds. For odd reasons, possibly to avoid institution duplication, when a reprint author is listed, the *Web of Science* removes the reprint institution from the list of represented institutions. As such, it is ambiguous which institution the first listed author is affiliated with, as he may be affiliated with the reprint institution *or* the first listed affiliation.

[8] See Table 4.4.

[9] When immunologists have dual affiliations, they will still almost always list their primary institution first, which is what I am primarily concerned with tracking.

[10] For example, if an immunologist moves from institution A to institution B and back to institution A. I only remove these moves if they occur in said succession. As another example, if an individual moves from institution A to institution B to institution C and subsequently to institution A, I record three moves: A→B, B→C, and C→D.

institution to the highest level. For example, *Harvard University, Biology Department* would aggregate up to *Harvard University.* As a result, inter-departmental but intra-institutional moves would not be considered a move. Third, and most importantly, I make use of string distance matching algorithms to ensure that two strings are distinct (a true move) and not simply typographical errors.[11]

Manually matching the output of the institution mapping and movement algorithms to immunology bios posted on www.isihighlycited.com[12] reveals strong overlap. In all cases, my movement algorithm does a very good job of identifying legitimate moves and never suffers from type I errors (false positives).

When an individual moves from one institution to a new one, a move instance occurs. As a result, my dichotomous dependent variable is 1 if immunologist i moves in year t and 0 otherwise.

4.3.3 Independent Variables

My two key independent variables are direct and indirect productivity. I use two variables to measure direct productivity: an immunologist's Impact Factor-weighted publications and his citations. Thomson ISI generates Impact Factor scores annually between 2000 and 2008. I retroactively weight all immunologist's publications by the journal's average Impact Factor in an attempt to capture the quality or impact of the publication. I collect forward citations from Thomson ISI's *Web of Science.*

I measure indirect productivity is measured as the number of acknowledgements received by the focal immunologist in the *Journal of Immunology.* As shown in Chapters 2 and 3, immunologists with more acknowledgements positively impact the performance of

---

[11] I rely on string distance algorithms as bundled through the FEBRL project (Christen, 2008). String distance algorithms calculate the "distance" between two strings, that is, how different two string are from one another. Larger distance implies greater dissimilarity. I use a combination of algorithms, including the Levenshtein and Longest Common Substring (LCS) (Gusfield, 1997).

[12] www.isihighlycited.com is a website run by Thomson ISI that provides biographies and curricula vitae of highly cited and prolific scientists in numerous natural and social sciences.

coauthors but are generally not fully financially compensated for this. I use the *Journal of Immunology* because it is both the second oldest immunology journal in the world and the most cited immunology journal. Adopting text name extraction techniques recently developed within the field of computer science (Councill, Giles, Han, and Manavoglu, 2005), I extract individual names from the acknowledgements section of all *Journal of Immunology* articles between 1960 and 1989 (inclusive). In these 30 years, 22,020 articles have been published, of which 13,950 contain at least one acknowledgement to an individual. I extract 53,133 acknowledgements instances in total from these 13,950 articles, yielding 2.4 acknowledgements per article written between 1960 and 1989.

4.3.4 Sample

The sample for this study is constrained by data availability. Immunologist mobility data are available from 1974 until 2007.[13] Direct productivity (publications and citations) data are available from 1960 until 2007, and indirect productivity data are available from 1960 until 1989. As such, I construct the sample using the intersection of available years: 1974-1989.

Three conditions must be met for an immunologist to enter the sample. First, the immunologist must have more than five lifetime immunology publications. I apply this restriction to ensure that all immunologists in the sample can be considered career scientists.[14] Second, the scientist must have published at least once between the time period of 1974 and 1989 so that I know that the immunologist is "at risk" of publishing during this time frame. Because not every immunologist publishes every year, I need to be able to differentiate between zeros that occur as a result of not producing an article in a given year versus a zero that occurs because the immunologist has yet to commence his career

---

[13] Even though immunologist affiliation data are available as of 1973, an additional year is required in order to determine if the affiliation of immunologist i is different in 1974 than in 1973, thus constituting a move.

[14] This also helps to remove graduate students who have been added to publications.

(or conversely has retired). Applying these first two restrictions results in a sample of 9,375 immunologists. Lastly, using the *Frequently Occurring First Names and Surnames From the 1990 Census*,[15] I remove all immunologists with surnames that occur 0.001 % (1 in 100,000) or more in the census population. This cut-off corresponds to eliminating 77.48% of the US population (in 1990) and represents the 18,839 most common surnames. I apply this restriction to ensure that I do not erroneously assume two distinct individuals to be the same person. Applying this restriction reduces the sample of immunologists from 9,375 to 4,665.

The final sample consists of the unbalanced panel of these 4,665 immunologists between 1974 and 1989, resulting in 35,126 immunologist-year observations.

4.3.5 Descriptive Statistics

Figure 4.1 presents raw trends on the count of both active immunologists and the number of move instances across time. Over the course of the sample period, both the number of publishing immunologists as well as the number of moves taking place in any given year appear to be increasing proportionately. So, while movement appears to be increasing over time, there do not appear to be more movers (as a percentage of immunologists) over time.

Figures 4.2 to 4.5 graphically demonstrate average publishing, Impact Factor-weighted publishing, citation, and acknowledgement rates, respectively, for immunologists who do and do not move in 1985.[16] Figure 4.2 plots publications for movers and non-movers across time. As can be seen, movers do appear to publish slightly more than non-movers. Figures 4.3 and 4.4 show similar patterns for Impact Factor-weighted publications and citations, respectively: Movers publish more in terms of both raw numbers and Impact Factor-weights and receive more citations than non-movers. Figure 4.5 looks at the

---

[15] http://www.census.gov/genealogy/names/

[16] The year 1985 is chosen randomly because it is temporally close to the middle of the sample.

annual acknowledgement rates for movers and non-movers. Here the reverse exists. Immunologists who move in 1985 have fewer acknowledgements in the two years prior than non-movers.

Table 4.1 compares movers to non-movers across the entire sample. Movers, on average, have more publications, Impact Factor-weighted publications, and citations than non-movers. Average annual acknowledgement levels, however, appear to be no different for movers than non-movers.

Taken as a whole, these descriptive statistics already appear to support the assertion that more observable (direct) productivity is associated with more mobility than unobservable (indirect) productivity.

Figures 4.6 and 4.7 presents histograms of the average experience (measured in years since the immunologist's first publication) of movers for their first move and for all moves, respectively. The average experience of immunologists for their first move is 5.7 years, and the average experience of immunologists at all moves is 6.9 years. As can be seen in Table 4.2, almost 65% of immunologists in the sample move only once, while the remaining 35% generate over half of all move instances (1,024 moves). Table 4.3 shows the top 10 institutions with respect to losing and receiving movers. Of the total 1,838 move instances, just under 13% of all move departures come from the top 10 institutions, while only 10% of movers go to a top 10 institution.

Lastly, Table 4.4 presents descriptive statistics for the entire sample. Comparing the large differences between the mean and median values reveals the heavy right tails of these distributions.

4.3.6 Econometric Estimation

Empirically, I am trying to demonstrate a relationship between immunologist mobility and his direct and indirect productivity in addition to the interaction between the two. Formally, one can view the relationship as follows:

$$Move_{it} = f(\beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + \beta_3 X_{it} \cdot Z_{it} + \varepsilon_{it}) \tag{4.5}$$

where the dependent variable $Move$ is 1 when immunologist i moves to a new institution in year t and 0 otherwise. X captures immunologist i's direct productivity (Impact Factor publications and citations) and Z his indirect productivity (acknowledgements). ε is a well-behaving error term. $\beta_1 > 0$ would provide support for Hypothesis 1, while $\beta_2 < 0$ would support Hypothesis 2.

For dichotomous dependent variables, logit or probit models are de rigueur. Unfortunately, interaction effects are difficult to interpret in nonlinear models (Ai and Norton, 2003). As such, I model Equation 4.5 as both a linear probability model (LPM) and as a logit function. In addition, because I am not interested in how the flow[17] of direct or indirect productivity impacts the probability of movement in the next year, I adopt a more flexible three-year moving average where all independent variables of interest are the sum of the covariates from two years prior through to and including the current year:

$$MovingAverage^3_{it} = \sum_{j=t-2}^{t} \zeta_{ij}$$

where ζ corresponds to one of the main independent variables: direct productivity, indirect productivity, and their interactions. All independent variables of interest in all empirical specifications are run with this three-year moving average unless otherwise noted. While movement data are only available after 1973, direct and indirect productivity data are available starting in 1960, thus allowing for the construction of these lagged moving averages.

I define the liner probability model (LPM) as:

$$\begin{aligned} \Pr(Y = 1|X) &= x'\beta \\ \mathbb{E}[Move_{it}|X, Z, \delta, \xi\, \phi] &= \beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + \beta_3 X_{it} \cdot Z_{it} + \delta_t + \xi_{it} + \phi_i + \varepsilon_{it} \quad (4.6) \end{aligned}$$

---

[17] Values in a given year.

where δ_t are year fixed effects, ξ_{it} are a set of 10 immunologist lifecycle experience dummies,[18] and ϕ_i are immunologist fixed effects. I estimate equation 4.6 using ordinary least squares (OLS). I use Huber-White robust standard errors with clustering at the immunologist level to correct the inherent heteroskedasticity in linear probability models and to allow for non-indepenendence across time.

I estimate the logit model as follows:

$$\begin{aligned}\Pr(Y=1|X) &= \frac{e^{x'\beta}}{1+e^{x'\beta}} = \Lambda(x'\beta) \\ \Pr(Move_{it}=1) &= \Lambda(\beta_0+\beta_1 X_{it}+\beta_2 Z_{it}+\beta_3 X_{it}\cdot Z_{it}+\delta_t+\xi_{it}+\phi_i) \qquad (4.7)\end{aligned}$$

Here too, δ_t and ϕ_i represent year and immunologist fixed effects, respectively. While I will directly estimate δ_t, ϕ_i, in practice, is conditioned out during estimation. I estimate the logit model by maximum likelihood and, as in the linear probability model, I report robust standard errors with immunologist clustering. Equations 4.6 and 4.7 serve as the main empirical specifications to be tested.

Concerns may exist over the relationship between direct productivity and mobility since due to the sampling strategy, I rely on publications to infer movement. As a result, my sample of movers may be more productive immunologists, *ceteris paribus*. I include the immunologist fixed effects to minimize this concern. By including immunologist fixed effects, I identify all mobility from within-immunologist variation in publishing and citation rates. In addition, as can be seen in Table 4.4, moves are not very strongly correlated with publications ($\rho = 0.08$). Lastly, I use the moving averages to minimize the impact of an immunologist's movement probability as a function of his publishing in any given year.

---

[18] These dummies attempt to semi-parametrically capture the immunologist's productivity over his lifecycle (Levin and Stephan, 1991). The 10 dummies correspond to his first 3, 6, 9, ... 30 years publishing in immunology. The omitted category is more than 30 years of experience in year t.

4.4 Results

The empirical objective of this study is to examine the relationship between a scientist's productivity and his mobility. In particular, I am interested in how an immunologist's observable and unobservable characteristics relate to inter-institutution mobility. Tables 4.5 and 4.6 present linear probability model results for pooled cross-sectional and fixed effects models, respectively. Table 4.7 presents both pooled logistical and conditional fixed effects logistical regression results to ensure that the LPM results are robust to an alternate functional form.

Column 1 presents evidence of the negative relationship between acknowledgements ($Acks$) and mobility but a positive relationship between Impact Factor-weighted publications ($IFpubs$) and mobility. Both signs are in line with what I have predicted. In addition, both relationships are statistically significant at the 1% level. These point estimates indicate that a one standard deviation increase in acknowledgements reduces the probability of moving by 0.7%, and a one standard deviation increase in an immunologist's Impact Factor-weighted publications corresponds to a 1.4% increase in the probability of moving, *ceteris paribus*.[19] Column 2 explores the second direct productivity measure of interest: citations ($Cites$). A one standard deviation increase in an immunologist's citations is associated with a 0.9% increase in the probability of mobility. Conversely, in this specification, a one standard deviation increase in acknowledgements is associated with a 0.5% decrease in the probability of mobility. Column 3 includes both $cites$ and $IFpubs$ in the same specification. Due to their high collinearity ($\rho = 0.73$), the results should be interpreted with caution. The negative impact of $Acks$ on mobility, however, continues to hold.

Column 4 introduces the interaction term between $Acks$ and $IFpubs$ into the specification. The negative sign on the interaction between these two variables ($Acks \cdot IFpubs$)

---

[19] Standard deviations of the variables can be seen in Table 4.4.

hints at direct and indirect productivity acting as substitutes for each other. Similar results appear in Column 5 where I interact *Acks* with *Cites*. However, due to the strong correlation between *Acks* and its interaction, the majority of the variation of *Acks* is absorbed, and thus the level point estimate of *Acks* in both Columns 4 and 5 is insignificant. I present Column 6 for the curious, but again, high multicollinearity results in estimates that are not necessarily unstable, but rather point estimates with large standard errors.

Table 4.6 replicates the specifications found in Table 4.5 but includes immunologist fixed effects. By including immunologist fixed effects, I control for all time invariant characteristics of the immunologist.[20] As a result, all identification of the parameters of interest come from within group (in this case the immunologist) and across-time variation in the covariates. While excluding all time-invariant factors that may influence an immunologist's decision to move is surely beneficial for statistical inference, the relationship between productivity and mobility is fairly stable across time, and so applying fixed effects can be seen as an exceptionally conservative test of the relationship between productivity and mobility as I do not make use of any between immunologist variation.

Column 1 in Table 4.6 shows a negative relationship between *Acks* and mobility and a statistically significant positive relationship between $IFpubs$ and mobility. The parameter estimate of *Acks* is more than half the size of the point estimate in Table 4.5 but still statistically significant at the 5% level. The parameter estimate for $IFpubs$ is very similar to the pooled cross-sectional estimate in Table 4.5, indicating that more within immunologist variation exists for publications than for acknowledgements. Indeed, 87% of the sample does not receive an acknowledgement over the course of the sample. Columns 4 and 5 introduce the interaction effects, which again are very similar in both magnitude and significance to the pooled cross-section estimates. In short, both the pooled cross-section and fixed effects linear probability models appear to provide support for hypotheses H1 and H2.

---

[20] These may include factors such as the immunologist's gender, schooling, prior experiences, etc.

One drawback to linear probability models is that the predicted values of the dependent variable may fall outside of the unit interval of 0 and 1, which would be theoretically impossible.[21] However, from the 35,126 observations analyzed in the linear probability models, only 65 fall outside the bounds of 0 and 1, allowing us to draw appropriate statistical inference from the point estimates (Wooldridge, 2002). Figures 4.8 and 4.9 show the distribution of predicted moves ($\widehat{Move}$) for Specification 1 from Table 4.5 (LPM) and Specification 1 from Table 4.7 (logit). Logit regressions alleviate this problem but also have many drawbacks, including difficulty in interpreting point estimates, in particular interactions (Ai and Norton, 2003). As such, while the linear probability models are preferred, I present the logit models to ensure consistency.

Table 4.7 presents pooled cross-sectional results in Columns 1 through 4, while I include immunologist fixed effects in Columns 5 through 8. Columns 1 and 2 replicate the specifications in Columns 1 and 2 of both linear probability model tables (Tables 4.5 and 4.6). I display marginal effects evaluated at the mean of all variables in square brackets below the standard errors. A comparison between the logit marginal effects in Table 4.7 and the point estimates shown in the linear probability model tables reveals a strong overlap in values: the marginal effect of *Acks* in Specification 1 is -0.0125, while the LPM point estimate for the same Specification in Table 4.5 is -0.0121. The similarity in results gives strong support to the use of LPM for modeling the main empirical specification (Equation 4.5). While high overlap between the logit and LPM models exists for the specification in Column 2, Columns 3 and 4, which incorporate the interactions, return less stable point estimates for *Acks*, which is incorrectly signed in Column 3 and insignificant in Column 4. Two reasons may account for this. First, the strong correlation between *Acks* and its interactions will cause the point estimates to be less stable. Second, as mentioned earlier and as shown in Ai and Norton (2003), interactions in non-

---

[21] When the predicted value $\hat{y} < 0$, then one can interpret there being a negative probability of mobility. Conversely, when $\hat{y} > 1$, the predicted probability of mobility occurring is over 100%.

linear models, such as the logit, can result in opposite signs and significance. This, in addition to the strong correlation with *Acks*, surely affects the estimation of the models. However, it should still be noted that the signs of the direct productivity measures (*IFpubs* and *Cites*) are identical and similar in magnitude to the LPM models. Lastly, Columns 5 through 8 include immunologist fixed effects. Because conditional fixed effects logit models cannot estimate parameters for observation groups without any variation in the dependent variable (either all 1s or all 0s), I conduct the analysis only on the 1,254 immunologists who move. The estimates in Columns 5 and 6 parallel the results reported so far in terms of signs, and all variables are significant at the 5% level except for *Acks* in Specification 6, which includes *Cites*.

In general, hypotheses 1 and 2 appear to hold across a number of specifications and function forms. I examine the robustness of these results in the next section.

4.4.1 Robustness Checks

Specifications in Tables 4.8 and 4.9 move away from the three-year moving average window that was used in Tables 4.5 through 4.7. Tables 4.8 adopts a six-year moving average defined as $MovingAverage^6_{it} = \sum_{j=t-5}^{t} \zeta_{ij}$, while Table 4.9 uses a cumulative stock of direct and indirect productivity measures, where $Stock_{it} = \sum_{j=0}^{t} \zeta_{ij}$.[22]

Going from a three-year moving average to a six-year moving average has minimal impact on the robustness of the results. While the parameter estimates in Table 4.8 are slightly smaller than those already reported, both the signage and significance of the variables of interest are unchanged. If anything, the results are slightly statistically stronger apart from the 10% statistical significance of *Acks* in Columns 3 and 4. While the significance of *Acks* disappears when I include immunologist fixed effects, significance is retained for all direct productivity measures in addition to their interactions at the 5% level or greater.

---

[22] $t = 0$ corresponds to the first year that immunologist i publishes (left truncated at 1960).

Table 4.9 reports results of linear probability regressions where the independent variables are stock variables instead of moving averages. Columns 1 through 4 again report pooled cross-sectional evidence and support Hypotheses 1 and 2, where all variables are significant at the 5% level or greater. Including immunologist fixed effects, however, greatly reduces the significance on all variables, as the fixed effects are all largely collinear with the stock variables. It is worth noting, however, that the signs of the parameter estimates are all in line with prior results. Taken as a whole, the results appear to not be sensitive to the moving average window used.

4.4.2 Supplementary Analysis

The analysis thus far has focused on the relationship between direct and indirect productivity and the dichotomous outcome of movement. While the theory developed in section 4.2.1 focuses on a scientist's movability, tacitly assuming that immunologists will only move if it is to their benefit, in reality this assumption is a bit strong. The mobility decision is not always endogenous to the scientist, and as such, I empirically explore the impact of direct and indirect productivity on what *types* of moves take place: to a higher-tiered institution, to an institution in the same tier, or to a lower-tiered institution.

Using annual publishing counts of all institutions, I am able to generate annual institution publishing distributions. To simplify analysis, I create four tiers. An institution is in the bottom tier if it falls below the 50th percentile of the distribution, in the second tier if it falls between the 50th and 75th percentile, in the third tier if it falls between the 75th and 90th percentile, and in the fourth (top) tier if it is above the 90th percentile. The three dependent variables used in Tables 4.10 to 4.13 reflect moves that change in tiers. I construct the variables as follows:

$Movedown$ = 1 if immunologist i moves down at least one tier from his current institution in year t, 0 otherwise.

$Movesame = 1$ if immunologist i moves to an institution in the same tier as his previous institution, 0 otherwise.

$Moveup = 1$ if immunologist i moves up at least one tier from his current institution in year t, 0 otherwise.

Table 4.10 presents the first results, whereby the *type* of move is a function of $Acks$ and $IFpubs$. $IFpubs$ appears to have an equal impact on mobility, irrespective of the tier the immunologist moves to. In a sense, this provides strong support for the main results shown. Namely, direct productivity has a positive impact on mobility for *all* types of moves. While the model developed does not provide insight into why more direct productivity would increase movement to a lower-tiered institution, intuitively we can think of instances in which institutions engage in department building by hiring the best (direct productivity) scholars in the field. When departments are attempting to improve their profile, they often hire scholars with the most visible productivity. $Acks$ on the other hand, has an increasingly negative impact on mobility as the immunologist moves up tiers. That is, immunologists with more acknowledgements are the least likely to move to a top-tier institution or, thought of differently, the immunologists who move up a tier are more likely to have low indirect productivity (fewer acknowledgements). This effect of moving up a tier holds even when I include immunologist fixed effects.

Table 4.11 presents very similar results to the previous table, but instead of $IFpubs$ serving as the direct productivity measure, I examine the impact of $Cites$ on mobility tier change. Again, across all specifications, direct productivity is positively associated with all mobility. Yet, when controlling for the immunologists citations, acknowledgements only have a negative association with mobility when the immunologist moves to a same-tier or higher-tiered institution. These results disappear, however, once I include immunologist fixed effects.

Tables 4.12 and 4.13 include interactions between $Acks$ and $IFpubs$ and $Cites$, respectively. In both tables, *Acks* only retains a statistically significant negative sign for moves to a higher-tiered institution. The positive relationship between direct productivity and moves holds across all specifications, while the negative relationship between the interaction of direct and indirect productivity holds for all pooled cross-sectional specifications and for most specifications with immunologist fixed effects.

In summary, while indirect productivity appears to have a differential impact on mobility depending on the change in tier of institutions, the effect does not hold for direct productivity. It may be that scientists with more indirect productivity are able to extract more value from lower-tiered schools. Alternatively, institutions may be more willing to make matching offers to immunologists with high indirect productivity from higher-tiered schools, resulting in less mobility. Again, moving to a lower-tiered institution may in and of itself not be a negative. A lower-tiered institution that is building up its immunology department may be a very enticing place to work for all types of immunologists.

4.5 Discussion and Conclusion

A robust correlation between productivity and mobility exists across a number of specifications and functional forms. Indirect productivity as measured by academic acknowledgements has a strong negative association with the probability of an immunologist moving, whereby a one standard deviation increase in an immunologist's indirect productivity is associated with a 0.6% decrease in mobility, on average. Direct productivity, as measured by Impact Factor-weighted publications and citation counts, when increased by one standard deviation, is associated with a 1% increase in the likelihood of mobility to a new institution.

These results have implications for multiple streams of literature. First, this study provides one of the first estimates of the relationship between different productivity types and mobility for academic scientists. A deeper understanding of the factors related to mobility is a necessary first step before examining the performance implications of labor mobility. Along this point, this study speaks to the literature on labor mobility-generated spillovers as it begins to open up the black box on the determinants of labor mobility, moving away from the assumption of homogenous human capital so often found in the literature. If the scientists providing the largest spillovers to their peers[23] (high indirect productivity scientists) are also the least mobile, the high spillovers we see from mobile scientists may be just the tip of the iceberg of the potential for human capital spillovers. Second, for public policy, academic science has a very strong impact on regional firm performance (Furman and MacGarvie, 2007). Consequently, a better model at predicting scientist mobility will allow us to move forward and begin to tease apart the differential effects of productivity and mobility on regional growth. Third and relatedly, a deeper understanding of human capital, in terms of both direct and indirect productivity and their role in transferring knowledge through mobility, is of chief concern not only to policy makers and strategy scholars but also practitioners. For example, what types of retention mechanisms can be put in place if the highest paid scientists (Chapter 3), are also the most mobile?

As can be seen from the relatively low R^2 values reported in the regression tables, the models appear to explain very little variation and the point estimates explain at most 1% of changes in annual mobility probabilities. Clearly the decision to move institutions is multifaceted, and one's observable and unobservable productivity are only two of many factors influencing both mobility opportunities and decisions. In addition, while immunologists with more direct productivity are indeed the most mobile, it may be that their home institutions provide matching offers, thus driving down the instances in which we actually observe mobility, a separate issue from how mobile an immunologist actually is. Nonetheless, while the results presented are robust to a number of specifications and

---

[23] As we saw in Chapter 2, more helpful scientists have a larger impact on peer performance than scientists with high (direct) productivity.

functional forms, these results should be seen as conservative initial estimates of the relationship between productivity and mobility.

Prior studies in entrepreneurship have shown a strong negative association between productivity and mobility (Agarwal, Ganco, and Ziedonis, 2007; Groysberg, Nanda, and Prats, 2007; Elfenbein, Hamilton, and Zenger, 2008). The common explanation for this empirical artifact is that individuals who derive more of their productivity from firm-specific skills are less likely to move. Just as private-sector knowledge workers develop firm-specific skills, the productivity of immunologists is also very institution specific as labs are highly immobile. The model developed in Section 4.2.1 provides a simple alternative explanation for low mobility rates of higher productivity individuals, whereby mobility is attenuated by how observable an individual's productivity is rather than its firm-specificity. As was demonstrated empirically, a joint increase in an immunologist's direct and indirect productivity results in lower mobility, a finding in line with prior research. Yet an individual's prime driver of mobility is still his direct (observable) productivity. Despite the importance of understanding the science of science, the generalizability of the empirics is less certain. While these cases do represent limitations to this study, optimistically they are also future research opportunities.

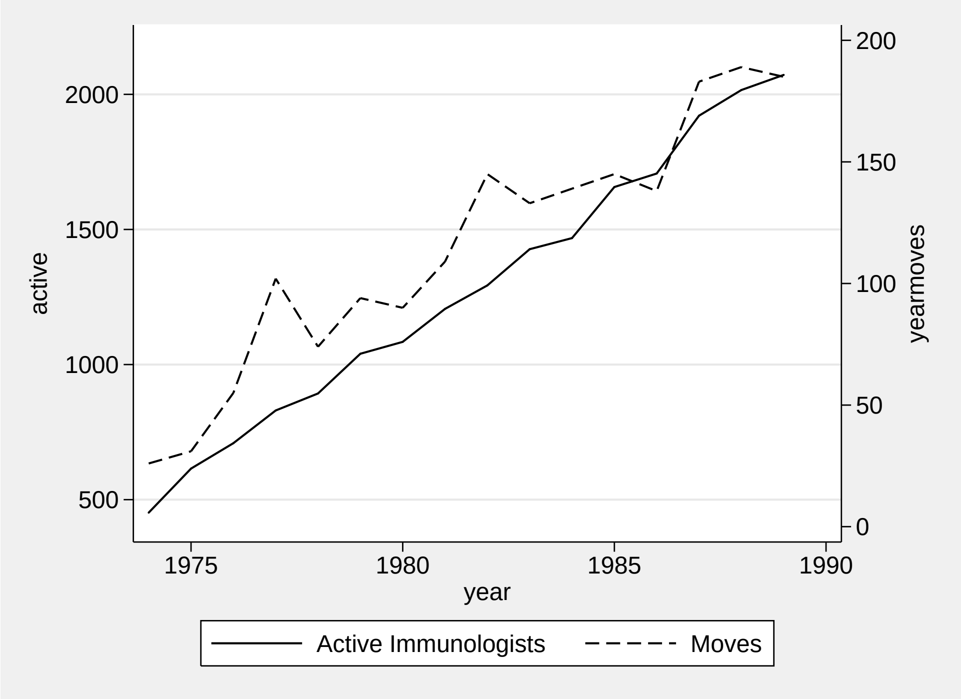

Figure 4.1: Immunologist Moves Over Time

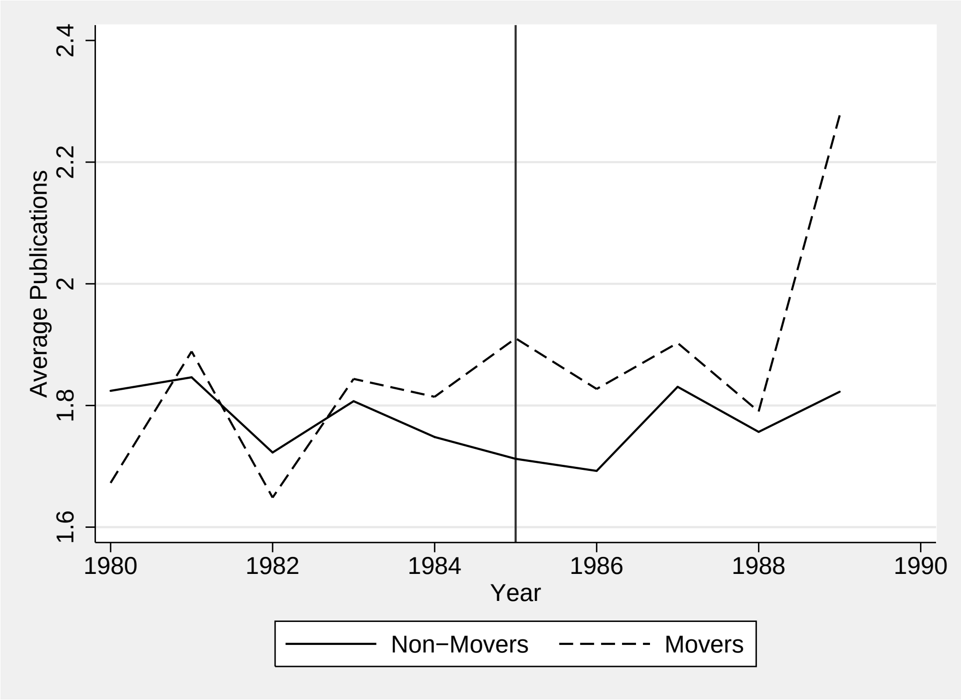

Figure 4.2: Publications Before and After 1985 Moves

Figure 4.3: Impact Factor Publications Before and After 1985 Moves

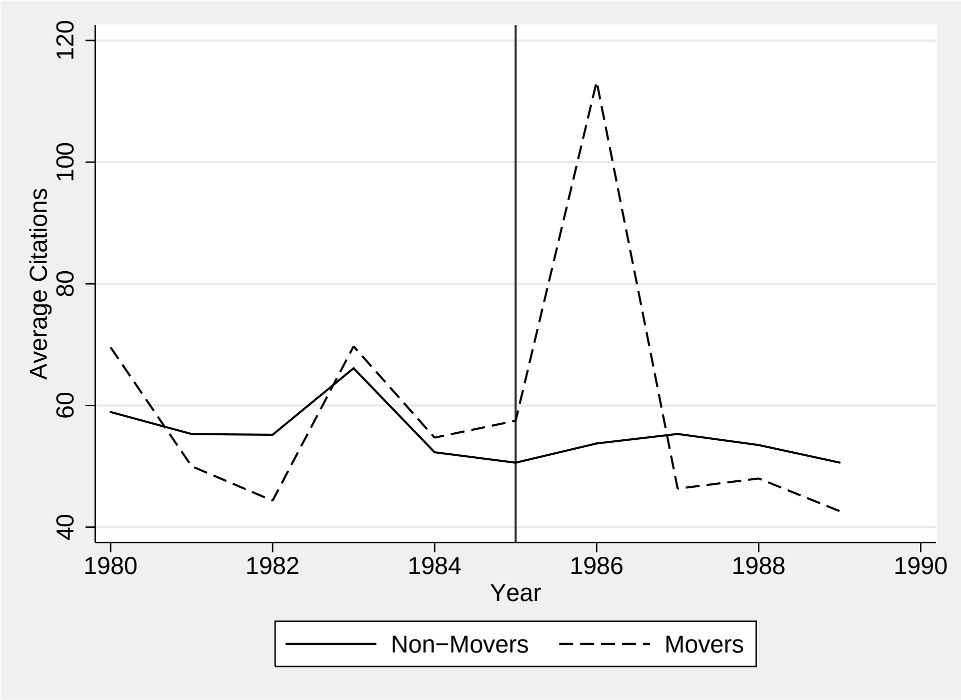

Figure 4.4: Citations Before and After 1985 Moves

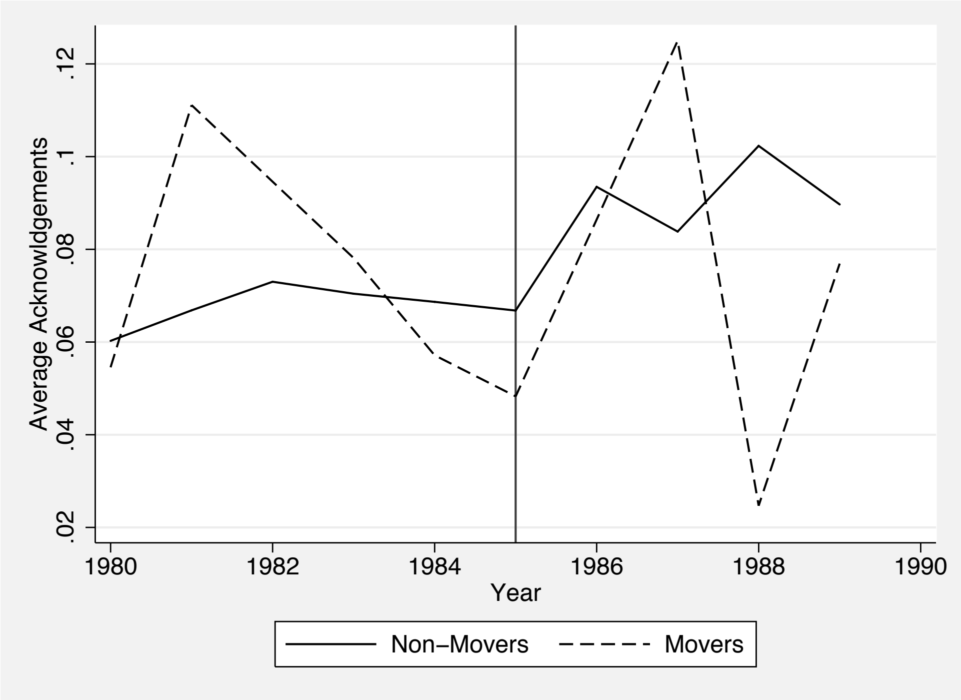

Figure 4.5: Acknowledgements Before and After 1985 Moves

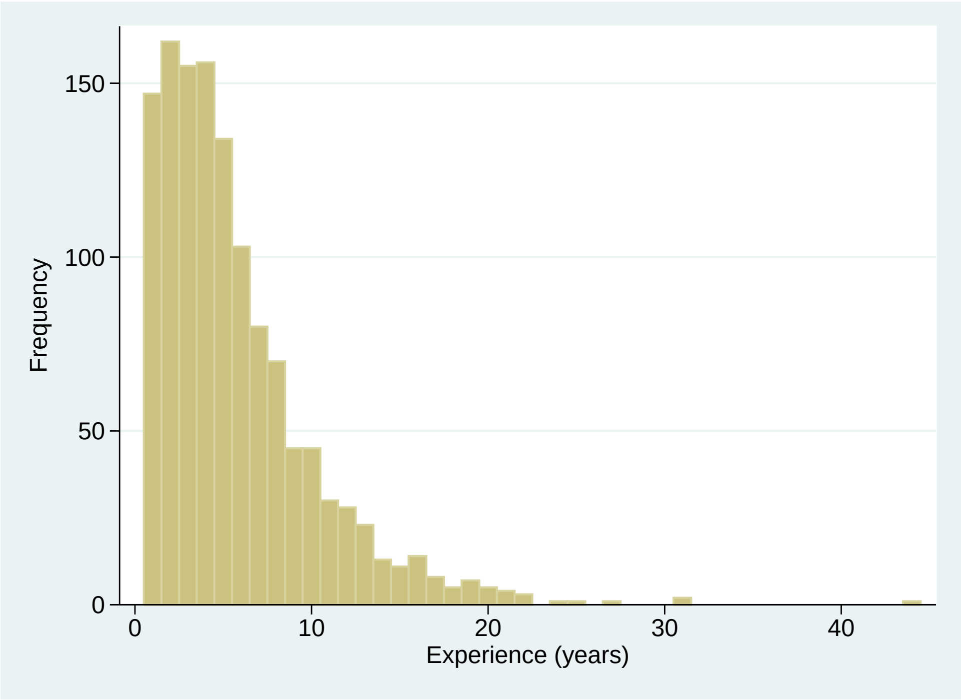

Figure 4.6: Immunologist Experience (in Years) at First Move

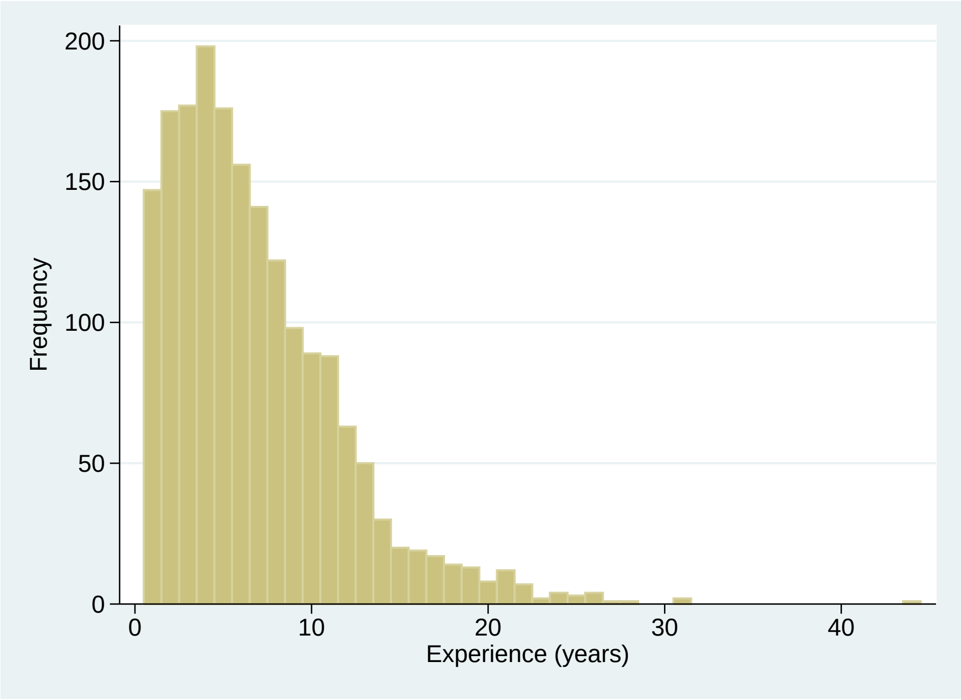

Figure 4.7: Immunologist Experience (in Years) at Time of Move

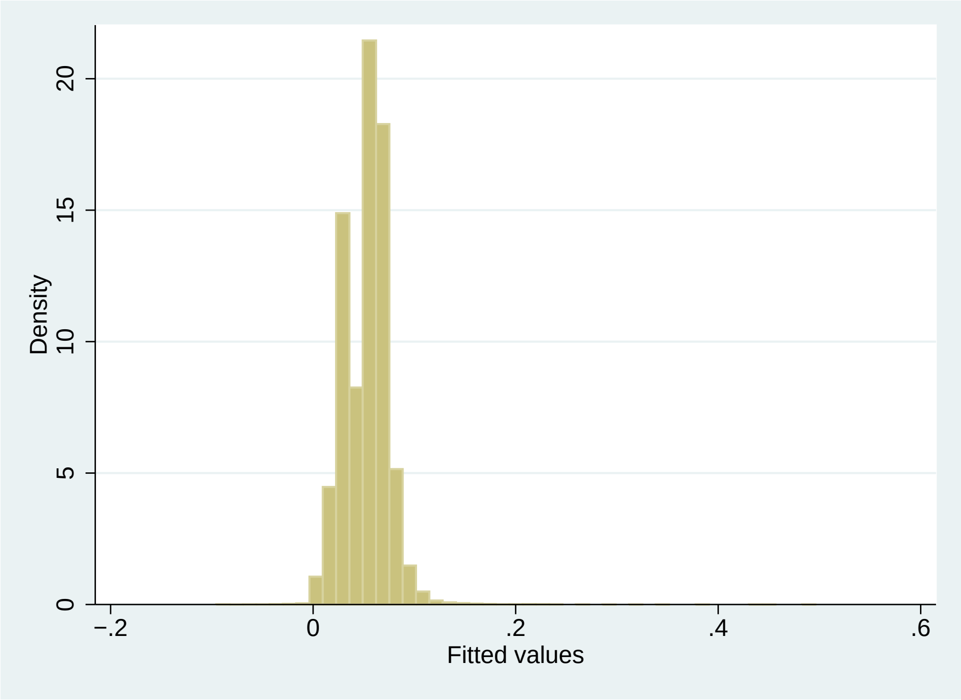

Figure 4.8: Histogram of LPM $\hat{y}$

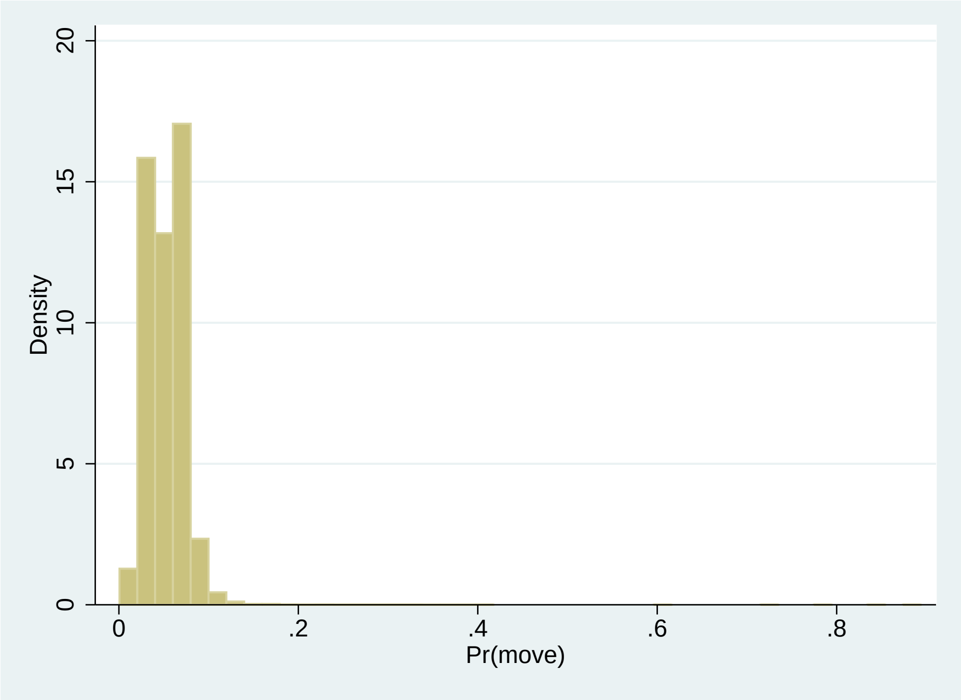

Figure 4.9: Histogram of Logit $\hat{y}$

Table 4.1: Descriptive Statistics: Movers vs. Non-Movers

	Mean	Std. Dev.	Min	Max
Full Sample				
N=35,126 Unique immunologists: 4,665 Average years: 7.53				
Publications	1.032	1.380	0	26.0
Impact Factor-Weighted Pubs	3.394	6.673	0	244.2
Citations	32.618	107.037	0	5793.0
Acknowledgements	0.042	0.281	0	12.0
Non-Movers				
N=22,073 Unique immunologists: 3,411 Average years: 6.47				
Publications	0.971	1.347	0	26.0
Impact Factor-Weighted Pubs	3.107	6.772	0	244.2
Citations	30.792	105.148	0	4412.0
Acknowledgements	0.042	0.294	0	12.0
Movers				
N=13,053 Unique immunologists: 1,254 Average years: 10.41				
Publications	1.136*	1.428	0	21.0
Impact Factor-Weighted Pubs	3.878*	6.473	0	96.6
Citations	35.705*	110.093	0	5793.0
Acknowledgements	0.042	0.258	0	10.0

* variables are statistically distinct from the Non-Movers sample at the 1% level.

Table 4.2: Number of Moves per Mover

Total Moves	Frequency	Percent	Cumulative Percent
1	814	64.91	64.91
2	326	26.00	90.91
3	90	7.18	98.09
4	18	1.44	99.52
5	6	0.48	100.00
Total Movers	1,254	100.00	
Total Moves	1,838		

Table 4.3: Top 10 Institutions by Movement Activity

Rank	Institution	Lost	Gained
1	HARVARD UNIV	28	28
2	ALL INDIA INST MED SCI	30	25
3	NCI	30	21
4	NIAID	29	20
5	SCRIPPS CLIN & RES FDN	22	16
6	UNIV TEXAS	16	20
7	OSAKA UNIV	20	15
8	UNIV TOKYO	20	15
9	POSTGRAD INST MED EDUC & RES	17	14
10	KYOTO UNIV	19	10
Total		231	184
Percentage of Total Moves		13%	10%

Table 4.4: Sample Descriptive Statistics and Correlations: N=35,126

	Variable	Mean	Median	Std. Dev.	Min	Max	1	2	3	4	5	6
1	$Move$	0.05	0	0.22	0	1.0						
2	$Pubs$	2.72	2.00	3.02	0	63.0	0.08					
3	$IFpubs$	9.07	4.76	15.79	0	601.1	0.05	0.76				
4	$Cites^*$	0.09	0.03	0.23	0	9.7	0.04	0.57	0.73			
5	$Acks$	0.11	0	0.60	0	24.0	0.00	0.28	0.42	0.36		
6	$Acks \cdot IFpubs$	5.04	0	86.07	0	6882.0	−0.01	0.38	0.58	0.46	0.64	
7	$Acks \cdot Cites$	0.06	0	1.16	0	81.1	−0.01	0.35	0.50	0.57	0.59	0.89

* Cites is scaled down to thousands and, as such, the mean value of .09 translates to 90 citations.

All variables except for Move are three-year moving averages: $t_{-2} \rightarrow t_0$.

Table 4.5: Main Results - Linear Probability Model (LPM) - Pooled Cross-Section

Dependent Variable:	$Move$	$Move$	$Move$	$Move$	$Move$	$Move$
	(1)	(2)	(3)	(4)	(5)	(6)
$Acks$	-0.0121^{**}	-0.0081^{**}	-0.0122^{**}	-0.0033	-0.0021	-0.0035
	(0.0020)	(0.0018)	(0.0019)	(0.0024)	(0.0020)	(0.0022)
$IFpubs$	0.0009^{**}		0.0008^{**}	0.0011^{**}		0.0011^{**}
	(0.0002)		(0.0003)	(0.0002)		(0.0002)
$Cites$		0.0407^{**}	0.0034		0.0551^{**}	0.0023
		(0.0125)	(0.0180)		(0.0131)	(0.0175)
$Acks \cdot IFpubs$				-0.0001^{**}		-0.0001^{**}
				(0.0000)		(0.0001)
$Acks \cdot Cites$					-0.0070^{**}	0.0017
					(0.0024)	(0.0020)
Intercept	0.0025	0.0075	0.0026	-0.0011	0.0055	-0.0010
	(0.0218)	(0.0221)	(0.0218)	(0.0219)	(0.0222)	(0.0218)
Experience FE	✓	✓	✓	✓	✓	✓
Observations	35126	35126	35126	35126	35126	35126
R^2	0.01	0.01	0.01	0.01	0.01	0.01

Standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.6: Main Results - Linear Probability Model (LPM) - Immunologist Fixed Effects

Dependent Variable:	$Move$	$Move$	$Move$	$Move$	$Move$	$Move$
	(1)	(2)	(3)	(4)	(5)	(6)
$Acks$	-0.0055^{*}	-0.0034	-0.0054^{+}	0.0036	0.0021	0.0036
	(0.0028)	(0.0027)	(0.0028)	(0.0035)	(0.0033)	(0.0036)
$IFpubs$	0.0010^{**}		0.0009^{**}	0.0013^{**}		0.0012^{**}
	(0.0002)		(0.0003)	(0.0002)		(0.0003)
$Cites$		0.0452^{**}	0.0136		0.0544^{**}	0.0117
		(0.0127)	(0.0163)		(0.0147)	(0.0178)
$Acks \cdot IFpubs$				-0.0001^{**}		-0.0001^{*}
				(0.0000)		(0.0000)
$Acks \cdot Cites$					-0.0056^{*}	-0.0009
					(0.0028)	(0.0020)
Intercept	-0.2572^{**}	-0.2724^{**}	-0.2567^{**}	-0.2525^{**}	-0.2715^{**}	-0.2523^{**}
	(0.0587)	(0.0584)	(0.0588)	(0.0588)	(0.0585)	(0.0589)
Experience FE	✓	✓	✓	✓	✓	✓
Immunologist FE	✓	✓	✓	✓	✓	✓
Observations	35126	35126	35126	35126	35126	35126
Groups	4665	4665	4665	4665	4665	4665
R^2	0.00	0.00	0.00	0.00	0.00	0.00

Standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.7: Main Results - Logistical Regression - Pooled Cross-Section and Immunologist Fixed Effects

Dependent Variable	*Move*	*Move*	*Move*	*Move*	*Move*	*Move*	*Move*	*Move*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
main								
Acks	-0.2727^{**}	-0.1636^{**}	0.1595^{**}	0.0844	-0.1509^{*}	−0.0949	0.3067^{**}	0.1668^{+}
	(0.0620)	(0.0495)	(0.0607)	(0.0583)	(0.0761)	(0.0732)	(0.1177)	(0.0967)
	[−0.0125]	[−0.0075]	[0.0071]	[0.0038]				
IFpubs	0.0111^{**}		0.0205^{**}		0.0220^{**}		0.0276^{**}	
	(0.0028)		(0.0028)		(0.0029)		(0.0031)	
	[0.0005]		[0.0009]					
Cites		0.4725^{**}		1.0571^{**}		0.7070^{**}		1.1058^{**}
		(0.1263)		(0.1530)		(0.1585)		(0.1843)
		[0.0218]		[0.0477]				
$Acks \cdot IFpubs$			-0.0128^{**}				-0.0139^{**}	
			(0.0025)				(0.0034)	
			[−0.0006]					
$Acks \cdot Cites$				-0.6951^{**}				-0.7114^{**}
				(0.1648)				(0.1959)
				[−0.0314]				
Intercept	-4.1268^{**}	-4.0734^{**}	-4.2319^{**}	-4.1346^{**}				
	(0.8486)	(0.8540)	(0.8406)	(0.8513)				
Experience FE	✓	✓	✓	✓	✓	✓	✓	✓
Immunologist FE					✓	✓	✓	✓
Observations	35126	35126	35126	35126	13053	13053	13053	13053
Groups					1254	1254	1254	1254
log likelihood	−7045.10	−7064.70	−7002.91	−7040.55	−3584.51	−3601.82	−3569.79	−3592.64
pseudo-R^2	0.02	0.02	0.03	0.02	0.04	0.03	0.04	0.04

Standard errors in parentheses.
Marginal effects (where applicable) in square brackets. The marginal effects retain the same significance as the point estimates.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.8: Robustness Checks: LPM - Six-Year Moving Average - Pooled Cross-Section and Fixed Effects

Dependent Variable	$Move$	$Move$	$Move$	$Move$	$Move$	$Move$	$Move$	$Move$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Acks$	-0.0081^{**}	-0.0056^{**}	-0.0032^{+}	-0.0026^{+}	-0.0042^{+}	-0.0031	0.0009	0.0003
	(0.0015)	(0.0013)	(0.0017)	(0.0014)	(0.0023)	(0.0022)	(0.0029)	(0.0027)
$IFpubs$	0.0004^{**}		0.0005^{**}		0.0004^{**}		0.0005^{**}	
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
$Cites$		0.0195^{**}		0.0280^{**}		0.0196^{**}		0.0258^{**}
		(0.0071)		(0.0077)		(0.0059)		(0.0065)
$Acks \cdot IFpubs$			-0.0000^{**}				-0.0000^{**}	
			(0.0000)				(0.0000)	
$Acks \cdot Cites$				-0.0018^{**}				-0.0013^{*}
				(0.0007)				(0.0006)
Intercept	0.0011	0.0075	-0.0029	0.0059	-0.2093^{**}	-0.2184^{**}	-0.2026^{**}	-0.2153^{**}
	(0.0216)	(0.0222)	(0.0214)	(0.0221)	(0.0680)	(0.0679)	(0.0680)	(0.0679)
Experience FE	✓	✓	✓	✓	✓	✓	✓	✓
Immunologist FE					✓	✓	✓	✓
Observations	35126	35126	35126	35126	35126	35126	35126	35126
Groups					4665	4665	4665	4665
R^2	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00

Standard errors in parentheses.
Independent variables are six-year moving averages defined as $MovingAverage^{6}_{it} = \sum_{j=t-5}^{t} \zeta_{ij}$.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.9: Robustness Checks: LPM - Stock Variables - Pooled Cross-Section and Fixed Effects

Dependent Variable	*Move*	*Move*	*Move*	*Move*	*Move*	*Move*	*Move*	*Move*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acks	-0.0059^{**}	-0.0044^{**}	-0.0039^{**}	-0.0033^{**}	-0.0042^{+}	-0.0051^{*}	-0.0042	-0.0050^{*}
	(0.0012)	(0.0011)	(0.0012)	(0.0010)	(0.0023)	(0.0022)	(0.0028)	(0.0026)
IFpubs	0.0002^{**}		0.0002^{**}		0.0000		0.0000	
	(0.0000)		(0.0001)		(0.0001)		(0.0001)	
Cites		0.0089^{*}		0.0127^{**}		0.0074		0.0076
		(0.0036)		(0.0046)		(0.0055)		(0.0062)
$Acks \cdot IFpubs$			-0.0000^{**}				-0.0000	
			(0.0000)				(0.0000)	
$Acks \cdot Cites$				-0.0004^{*}				-0.0000
				(0.0002)				(0.0002)
Intercept	-0.0101	0.0064	-0.0189	0.0033	-0.2279^{**}	-0.2267^{**}	-0.2279^{**}	-0.2267^{**}
	(0.0199)	(0.0231)	(0.0187)	(0.0224)	(0.0678)	(0.0678)	(0.0678)	(0.0678)
Experience FE	✓	✓	✓	✓	✓	✓	✓	✓
Immunologist FE					✓	✓	✓	✓
Observations	35126	35126	35126	35126	35126	35126	35126	35126
Groups					4665	4665	4665	4665
R^2	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00

Standard errors in parentheses.

Independent variables are cumulative stocks defined as $Stock_{it} = \sum_{j=0}^{t} \zeta_{ij}$, where $t = 0$ is the first year that immunologist i publishes.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.10: Supplementary Analysis: LPM - Move Type - $IFpubs$

Dependent Variable	$Movedown$	$Movedown$	$Movesame$	$Movesame$	$Moveup$	$Moveup$
	(1)	(2)	(3)	(4)	(5)	(6)
$Acks$	-0.0022^{*}	-0.0001	-0.0044^{**}	-0.0025	-0.0055^{**}	-0.0028^{*}
	(0.0010)	(0.0013)	(0.0011)	(0.0016)	(0.0010)	(0.0014)
$IFpubs$	0.0002^{**}	0.0002^{**}	0.0003^{**}	0.0004^{**}	0.0003^{**}	0.0004^{**}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Intercept	-0.0132^{**}	-0.1182^{**}	-0.0066^{+}	-0.0572	0.0223	-0.0818^{*}
	(0.0020)	(0.0316)	(0.0039)	(0.0358)	(0.0219)	(0.0360)
Experience FE	✓	✓	✓	✓	✓	✓
Immunologist FE		✓		✓		✓
Observations	35126	35126	35126	35126	35126	35126
Groups		4665		4665		4665
R^2	0.00	0.00	0.00	0.00	0.00	0.00

Standard errors in parentheses.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.11: Supplementary Analysis: LPM - Move Type - *Cites*

Dependent Variable	*Movedown*	*Movedown*	*Movesame*	*Movesame*	*Moveup*	*Moveup*
	(1)	(2)	(3)	(4)	(5)	(6)
Acks	-0.0017	0.0002	-0.0027^{**}	-0.0016	-0.0037^{**}	-0.0020
	(0.0010)	(0.0013)	(0.0010)	(0.0015)	(0.0009)	(0.0014)
Cites	0.0131^{*}	0.0168^{*}	0.0149^{**}	0.0123^{*}	0.0127^{**}	0.0162^{**}
	(0.0057)	(0.0076)	(0.0053)	(0.0055)	(0.0044)	(0.0054)
Intercept	-0.0122^{**}	-0.1197^{**}	-0.0046	-0.0644^{+}	0.0242	-0.0882^{*}
	(0.0019)	(0.0315)	(0.0037)	(0.0358)	(0.0221)	(0.0358)
Experience FE	✓	✓	✓	✓	✓	✓
Immunologist FE		✓		✓		✓
Observations	35126	35126	35126	35126	35126	35126
Groups		4665		4665		4665
R^2	0.00	0.00	0.00	0.00	0.00	0.00

Standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.12: Supplementary Analysis: LPM - Move Type - $Acks \cdot IFpubs$

Dependent Variable	$Movedown$	$Movedown$	$Movesame$	$Movesame$	$Moveup$	$Moveup$
	(1)	(2)	(3)	(4)	(5)	(6)
$Acks$	0.0004	0.0023	-0.0010	0.0010	-0.0027^{*}	0.0004
	(0.0012)	(0.0018)	(0.0014)	(0.0021)	(0.0012)	(0.0018)
$IFpubs$	0.0003^{**}	0.0003^{**}	0.0004^{**}	0.0005^{**}	0.0004^{**}	0.0005^{**}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$Acks \cdot IFpubs$	-0.0000^{**}	-0.0000^{**}	-0.0001^{**}	-0.0000^{*}	-0.0000^{*}	-0.0000^{*}
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	-0.0142^{**}	-0.1170^{**}	-0.0080^{*}	-0.0555	0.0211	-0.0801^{*}
	(0.0022)	(0.0316)	(0.0040)	(0.0359)	(0.0219)	(0.0361)
Experience FE	✓	✓	✓	✓	✓	✓
Immunologist FE		✓		✓		✓
Observations	35126	35126	35126	35126	35126	35126
Groups		4665		4665		4665
R^2	0.00	0.00	0.00	0.00	0.00	0.00

Standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.13: Supplementary Analysis: LPM - Move Type - $Acks \cdot Cites$

Dependent Variable	*Movedown*	*Movedown*	*Movesame*	*Movesame*	*Moveup*	*Moveup*
	(1)	(2)	(3)	(4)	(5)	(6)
Acks	0.0002	0.0016	−0.0004	0.0004	-0.0019^{+}	0.0001
	(0.0011)	(0.0017)	(0.0012)	(0.0020)	(0.0010)	(0.0017)
Cites	0.0176^{**}	0.0193^{*}	0.0205^{**}	0.0155^{*}	0.0171^{**}	0.0196^{**}
	(0.0060)	(0.0081)	(0.0058)	(0.0067)	(0.0052)	(0.0065)
$Acks \cdot Cites$	-0.0022^{**}	-0.0015^{*}	-0.0027^{**}	−0.0020	-0.0021^{*}	−0.0021
	(0.0007)	(0.0007)	(0.0010)	(0.0013)	(0.0009)	(0.0013)
Intercept	-0.0128^{**}	-0.1195^{**}	−0.0053	-0.0641^{+}	0.0236	-0.0879^{*}
	(0.0020)	(0.0315)	(0.0037)	(0.0359)	(0.0221)	(0.0359)
Experience FE	✓	✓	✓	✓	✓	✓
Immunologist FE		✓		✓		✓
Observations	35126	35126	35126	35126	35126	35126
Groups		4665		4665		4665
R^2	0.00	0.00	0.00	0.00	0.00	0.00

Standard errors in parentheses.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Bibliography

Acemoglu, D. (1996): "A Microfoundation for Social Increasing Returns in Human Capital Accumulation," *Quarterly Journal of Economics*, 111(3), 779–804.

Acemoglu, D., and J. Linn (2004): "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, 119(3), 1049–1090.

Agarwal, R., M. Ganco, and R. H. Ziedonis (2007): "Reputations for Toughness in Patent Enforcement: Implications for Knowledge Spillovers via Inventor Mobility," Mimeo, University of Illinois.

Agrawal, A., I. Cockburn, and J. McHale (2006): "Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships," *Journal of Economics Geography*, 6(5), 571–591.

Agrawal, A., D. Kapur, and J. McHale (2008): "How do spatial and social proximity influence knowledge flows? Evidence from patent data," *Journal of Urban Economics*, 64(2), 258–269.

Ai, C., and E. C. Norton (2003): "Interaction Terms in Logit and Probit Models," *Economics Letters*, 80, 123–129.

Allison, P. D., and J. S. Long (1990): "Departmental Effects on Scientific Productivity," *American Sociological Review*, 55(4), 469–478.

Almeida, P., and B. Kogut (1999): "Localization of Knowledge and the Mobility of Engineers in Regional Networks," *Management Science*, 45(7), 905–917.

Arrow, K. J. (1962): "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. by R. R. Nelson, National Bureau of Economic Research, Conference Series. Princeton University Press, Princeton.

Audretsch, D. B., and M. P. Feldman (1996): "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86(3), 630–640.

Azariadis, C., and A. Drazen (1990): "Threshold Externalities in Economic Development," *The Quarterly Journal of Economics*, 105(2), 501–526.

Azoulay, P., J. Graff Zivin, and J. Wang (2008): "Superstar Extinction," Working Paper.

Barney, J. (1986): "Strategic Factor Markets," *Management Science*, 32, 1231–1241.

——— (1991): "Firm Resources and Sustained Competitive Advantage," *Journal of Management*, 17(1), 99–120.

Becker, G. S. (1962): "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 70(s5), 9.

Bennett, B. (1965): "Specific Suppression of Tumor Growth By Isolated Peritioneal Macrophages from Immunized Mice," *Journal of Immunology*, 95(4), 656–664.

Brown, J. (2008): "Quitter Never Win: The (Adverse) Incentive Effects of Competing with Superstars," Working Paper.

Christen, P. (2008): "Febrl - A Freely Available Record Linkage System with a Graphical User Interface," in *Second Australasian Workshop on Health Data and Knowledge Management (HDKM 2008)*, ed. by J. R. Warren, P. Yu, J. Yearwood, and J. D. Patrick, vol. 80 of *CRPIT*, pp. 17–25, Wollongong, NSW, Australia. ACS.

Cockburn, I. M., and R. M. Henderson (1998): "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery," *Journal of Industrial Economics*, 46(2), 157–182.

Councill, I. G., C. L. Giles, H. Han, and E. Manavoglu (2005): "Automatic acknowledgement indexing: expanding the semantics of contribution in the CiteSeer digital library," in *K-CAP '05: Proceedings of the 3rd International Conference on Knowledge Capture*, pp. 19–26, New York, NY, USA. ACM.

Elfenbein, D. W., B. H. Hamilton, and T. R. Zenger (2008): "The Entrepreneurial Spawning of Scientists and Engineers: Stars, Slugs, and the Small Firm Effect," Mimeo, Washington University in St. Louis.

Ernst, H., C. Leptien, and J. Vitt (2000): "Inventors are not alike: The distribution of patenting output among industrial R&D personnel," *IEEE Transactions on Engineering Management*, 47(2), 184–199.

Fleming, L., S. Mingo, and D. Chen (2007): "Collaborative Brokerage, Generative Creativity, and Creative Success," *Administrative Science Quarterly*, 52(3), 443–475.

Fletcher, G. J. O., and M. S. Clark (2003): *Blackwell Handbook of Social Psychology: Interpersonal Processes*. Blackwell Publishing, Malden, MA.

Furman, J. L., and M. J. MacGarvie (2007): "Academic Science and the Birth of Industrial Research Laboratories in the U.S. Pharmaceutical Industry," *Journal of Economic Behavior & Organization*, 63(4), 756 – 776, Academic Science and Entrepreneurship: Dual engines of growth.

Gandossy, R. P., E. Tucker, and N. Verma (2006): *Workforce Wake-Up Call: Your Workforce is Changing, Are You?* John Wiley and Sons.

Gilfillan, S. C. (1935): *Inventing the Ship*. Follett Publishing Company, Chicago.

Groysberg, B., L.-E. Lee, and A. Nanda (2008): "Can They Take It With Them? The Portability of Star Knowledge Workers' Performance," *Management Science*, 54(7), 1213–1230.

Groysberg, B., A. Nanda, and M. J. Prats (2007): "Does Individual Performance Affect Entrepreneurial Mobility? Empirical Evidence From The Financial Analysis Market," NBER Working Paper #13633.

Gusfield, D. (1997): *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge University Press.

Guthridge, M., A. B. Komm, and E. Lawson (2008): "Making Talent a Strategic Priority," *McKinsey Quarterly*, 1, 48–59.

HACKETT, C. J., D. ROTROSEN, H. AUCHINCLOSS, AND A. S. FAUCI (2007): "Immunology research: challenges and opportunities in a time of budgetary constraint," *Nature Immunology*, 8(2), 114–117.

HAUSMAN, J., B. H. HALL, AND Z. GRILICHES (1984): "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica*, 52(4), 909–938.

HIRSHLEIFER, J., A. GLAZER, AND D. HIRSHLEIFER (1998): *Price Theory and Applications*. Prentice Hall, Upper Saddle River, NJ.

HOISL, K. (2007): "Tracing mobile inventors – The causality between inventor mobility and inventor productivity," *Research Policy*, 36(5), 619 – 636.

JACOBS, J. (1969): *The Economy of Cities*. Random House, New York, NY.

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), 577–598.

JONES, B. F. (2009): "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?," *Review of Economic Studies*, 76(1), 283–317.

JOVANOVIC, B. (1979): "Job Matching and the Theory of Turnover," *The Journal of Political Economy*, 87(5), 972–990.

JOVANOVIC, B., AND R. MOFFITT (1990): "An Estimate of a Sectoral Model of Labor Mobility," *The Journal of Political Economy*, 98(4), 827–852.

KAPUR, D., AND J. MCHALE (2005): *Give Us Your Best and Brightest*. Center for Global Development, Washington DC.

KOGUT, B., AND U. ZANDER (1992): "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology," *Organization Science*, 3(3), 383–397.

LAWRENCE, H. S., AND M. COHN (1993): "In Memoriam: Dr. Maurice Landy," *Cellular Immunology*, 152, 1–6.

LAZEAR, E. P. (1986): "Raids and Offer Matching," *Research in Labor Economics*, 8, 141–165.

Levin, S. G., and P. E. Stephan (1991): "Research Productivity Over the Life Cycle: Evidence for Academic Scientists," *The American Economic Review*, 81(1), 114–132.

Lotka, A. J. (1926): "The Frequency Distribution of Scientific Productivity," *Journal of the Washington Academy of Science*, 16, 317–325.

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

MacDonald, G. M. (1988): "Job Mobility in Market Equilibrium," *The Review of Economic Studies*, 55(1), 153–168.

Makadok, R. (2001): "Toward a Synthesis of the Resource-Based and Dynamic-Capability Views of Rent Creation," *Strategic Management Journal*, 22, 387–401.

Marx, M., D. Strumsky, and L. Fleming (2009): "Mobility, Skills, and the Michigan Non-compete Experiment," *Management Science*, 55(6), 875–889.

Merton, R. K. (1973): *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Møen, J. (2005): "Is Mobility of Technical Personnel a Source of R&D Spillovers?," *Journal of Labor Economics*, 23(1), 81–114.

Narin, F., and A. Breitzman (1995): "Inventive Productivity," *Research Policy*, 24(4), 507–519.

Nelson, R. R., and S. G. Winter (1982): *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press, Cambridge, MA.

Oettl, A., and A. Agrawal (2008): "International labor mobility and knowledge flow externalities," *Journal of International Business Studies*, 39(8), 1242–1260.

Pakes, A., and Z. Griliches (1980): "Patents and R&D at the firm level: A first report," *Economics Letters*, 5(4), 377–381.

Peteraf, M. A. (1993): "The Cornerstones of Competitive Advantage: A Resource-Based View," *Strategic Management Journal*, 14(3), 179–191.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5), S71–S102.

Rosen, S. (1981): "The Economics of Superstars," *The American Economic Review*, 71(5), 845–858.

Rosenkopf, L., and P. Almeida (2003): "Overcoming Local Search through Alliances and Mobility," *Management Science*, 49(6), 751–766.

Schankerman, M., R. Shalem, and M. Trajtenberg (2006): "Software Patents, Inventors and Mobility," Mimeo, London School of Economics.

Silverstein, A. M., and B. Benacerrafe (2001): "In Memoriam: Philip Gell," *Cellular Immunology*, 213, 1–3.

Singh, J. (2005): "Collaborative Networks as Determinants of Knowledge Diffusion Patterns," *Management Science*, 51, 756–770.

Smith, C. A., D. W. Organ, and J. P. Near (1983): "Organizational Citizenship Behavior: Its Nature and Antecedents," *Journal of Applied Psychology*, 68(4), 653–663.

Song, J., P. Almeida, and G. Wu (2003): "Learning-by-Hiring: When Is Mobility More Likely to Facilitate Interfirm Knowledge Transfer?," *Management Science*, 49(4), 351–365.

Steinman, R. M., and C. L. Moberg (1994): "Zanvil Alexander Cohn 1926-1993," *Journal of Experimental Medicine*, 179(1), 1–30.

Stephan, P. E. (1996): "The Economics of Science," *Journal of Economic Literature*, 34(3), 1199–1235.

Stern, S. (2004): "Do Scientists Pay to Be Scientists?," *Management Science*, 50(6), 835–853.

Tam, P.-W., and K. J. Delaney (2005): "Talent Search: Google's Growth Helps Ignite Silicon Valley Hiring Frenzy," *The Wall Street Journal*, 23 November 2005, A1.

Tedder, T. F., and J. R. Dawson (2003): "In Memoriam: D. Bernard Amos," *Journal of Immunology*, pp. 6316–6317.

Tirole, J. (1993): *The Theory of Industrial Organization.* MIT Press, Cambridge, MA.

Topel, R. (1991): "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *The Journal of Political Economy*, 99(1), 145–176.

TOPEL, R. H., AND M. P. WARD (1992): "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 107(2), 439–479.

WALDINGER, F. (2008): "Peer Effects in Science – Evidence from the Dismissal of Scientists in Nazi Germany," Mimeo, London School of Economics.

WERNERFELT, B. (1984): "A Resource-Based View of the Firm," *Strategic Management Journal*, 5(2), 171–180.

WOOLDRIDGE, J. M. (1999): "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics*, 90(1), 77–97.

——— (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA.

WUCHTY, S., B. F. JONES, AND B. UZZI (2007): "The Increasing Dominance of Teams in Production of Knowledge," *Science*, 316, 1036–1039.

YUNIS, E. J. (2004): "D. Bernard Amos," in *Biographical Memoirs*, ed. by N. A. o. S. Office of the Home Secretary, vol. 85, pp. 1–19. The National Academies Press.

ZELLNER, C. (2003): "The economic effects of basic research: evidence for embodied knowledge transfer via scientists' migration," *Research Policy*, 32(10), 1881 – 1895.

ZUCKER, L. G., M. R. DARBY, AND M. B. BREWER (1998): "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *The American Economic Review*, 88(1), 290–306.

ZUCKER, L. G., M. R. DARBY, AND M. TORERO (2002): "Labor Mobility from Academe to Commerce," *Journal of Labor Economics*, 20(3), 629–660.