Exploring the Research Behavior of University Professors
Through Patent Data

Manish G. Patel
Department of Economics
Stanford University, Stanford, CA 94305-6072 USA
Email: manishpatel@stanford.edu

Abstract

This paper examines the behavior of academic researchers within the context of university-industry relationships. Two aspects of a researcher’s behavior are explored: 1) the environmental effects of institution and sponsorship on research characteristics and 2) the anticipatory career behavior of a researcher. To investigate these issues we develop a formal utility model for a researcher. Using individual-level patent portfolio data for Stanford University researchers, we empirically test our model using categorical data analysis techniques. We find that environmental effects do not impact the appropriability of a researcher’s work but anticipatory career behavior does. These findings imply industry relationships with universities do not influence a researcher’s behavior and the migration of researchers from universities to industry can facilitate knowledge transfer.
Acknowledgements

Foremost I would like thank Professor Tim Bresnahan for his extraordinary guidance, insight, and support. My thesis would not have been possible without his commitment throughout the many stages of this project. In addition, I want to thank Professor Geoff Rothwell for his excellent advice and humor throughout this process. Also, I am grateful to Professor Brent Goldfarb, Professor Arvids Ziedonis, and Kathy Ku and Rob Scholes of the Stanford Office of Technology Licensing for their time and assistance in helping me obtain my data.
1 INTRODUCTION

Over the last twenty years there has been increasing interest in the relationship between American universities and industry. A tremendous amount of this attention has focused on how universities can foster technical advance in industry. As interest in the university-industry relationship grows, so grows the economic literature exploring this complex relationship. Within the literature, an emphasis has been placed on interactions at the institutional level and the channels of interaction such as patents and publications. The incentives university researchers face and their subsequent behavior, however, has received less attention. This study seeks to add to the growing body of university-industry literature by developing a behavioral model of university researchers. We investigate two questions in this paper. First, what is the effect of different institutional incentive structures on the behavior of a researcher? Second, does an academic researcher exhibit anticipatory career behavior due to future incentives?

The interaction of American universities and industry occurs at a number of levels that vary from minimal funding support for a project to long-term joint collaboration on an endeavor. Fundamentally, the university researcher is the conduit through which the academic environment interacts with the commercial environment. In addition to working with their commercial counterparts, university professors also become these counterparts. The migration of university researchers from academia to industry is another important form of interaction that has received little attention. During these different interactions, two important issues arise: first, improving knowledge transfer and technical advances from universities to industry and second, preventing the commercialization of academic values. Since the university professor is central to both of these issues, a richer understanding of them can be gained by studying her behavior.
The knowledge transfer from university to industry is an important aspect of technical advance in American industry. Throughout American industrial history, a significant portion of breakthroughs and innovations have resulted from corporate relationships with universities. As university-industry relationships grow closer these knowledge transfers are a more frequent occurrence. By understanding the behavior of a university professor, the dynamics of these transfers can be better understood. In particular, understanding the behavior of a university professor may illuminate other forms of knowledge transfer outside of the traditional channels of patents, publications, and conferences.

As knowledge transfers take place between universities and industry, the commercialization of academic research becomes an important consideration. The proponents of university-industry relationship argue that such interactions can only facilitate the primary objective of a research institution, which is to develop a rich body of knowledge around a subject. In contrast, critics argue that industry involvement with universities corrupts the academic researcher by encouraging her to deviate from creating pure knowledge to developing applied knowledge.

This paper’s principal purpose is to develop an understanding of the behavior of university professors to enhance our understanding of knowledge transfers and the commercialization of academic values. To investigate the behavior of professors, we employ a unique approach that analyzes patent and sponsorship data from Stanford University. The next section provides a background on related university-industry literature and discusses techniques for using patent data. Section 3 develops a theoretical behavioral model for university researchers and presents the hypotheses tested. Section 4 introduces the data used in the
empirical analysis of the behavioral model. Section 5 presents the results and analyses of our hypothesis testing. Section 6 explores the implications of our findings and concludes.

2 REVIEW OF LITERATURE

2.1 Background

Gradually, over the last 40 years there has been an increase in the interaction between universities and industry. This interaction has played a vital role in the technical advance of American industry (Rosenberg and Nelson, 1994). The complex relationships between universities and industry have begun to attract increasing scrutiny because of the growing intensity of these interactions. Within the economic literature, four distinct categories of research have emerged that explore the intricacies of the university-industry relationship (Agrawal, 2001). The four categories are firm characteristics, university characteristics, geography of knowledge spillovers, and channels of knowledge transfer. Firm characteristics constitute factors of companies that allow them to successfully interact with universities and take advantage of external research endeavors (Cohen and Levinthal 1989, Cohen and Levinthal 1990, Cockburn and Henderson 1998). Alternatively, research in the university characteristics category emphasizes issues that relate more directly to universities such as understanding licensing revenue systems, intellectual property protection strategies, and incentives for academic researchers (Audrestch and Stephan 1996, Dasgupta and David 1994, Henderson, Jaffe, and Trajtenberg 1998a, Henderson, Jaffe, and Trajtenberg 1998b, Jensen and Thursby 1998, Stern 1999). Literature on the geography of knowledge spillovers investigates spatial relationships between private firms and universities, evaluating performance of firms with respect to university knowledge spillovers (Audrestch and Feldman 1996, Feldman 1994, Jaffe 1989, Jaffe, Trajtenberg, Henderson 1993, Varga 1998). The final category, channels of
knowledge transfer, explores the differences among conduits of knowledge sharing between universities and industry such as publications and patents (Bania, Eberts, and Fogarty 1993, Agrawal and Henderson 2002, Shane 2002). These categories obviously have a great deal of interrelatedness and are influenced by one another. Of particular relevance to this paper is the growing research in the area of university characteristics as it relates to channels of knowledge transfer.

The relationship between American universities and industry began to flourish in the early 1980s. Newberg and Dunn (2002) explain that in response to a variety of competitive pressures companies reduced their internal R&D programs. Simultaneously, however, advanced technologies became a critical competitive factor for many industries. As a result, companies sought out various collaborative endeavors with other firms as well as with universities to meet their R&D requirements.

Newberg and Dunn (2002) uncover the major benefits to both universities and industry by examining the interactions of these institutions over the last 20 years. For universities, four primary benefits emerge from their relationships with industry: “1) Access to industry resources including financial support and advanced technology, 2) superior training and placement opportunities for students, 3) the stimulation and exposure to current industry problems, and 4) the income from commercially valuable inventions” (p. 194). The third point, exposure to industry problems, has been a major aspect of many American universities and is explored in greater depth in Rosenberg and Nelson (1994). The primary benefits to industry are access to advance academic research and prestige. Secondarily, corporate firms can aggressively position themselves to recruit highly-qualified students.
Critics of university-industry relationships de-emphasize the benefits to both sides of the interaction. They argue that the university-industry relationship is “A threat to the integrity of academic research. They despair that greater involvement with industry and commerce will corrupt academic research and teaching, divert attention away from fundamental research, and potentially destroy the openness of communication among university scientists that is such an essential component of academic research” (Rosenberg and Nelson, 1994, p. 323-4). On the other side of the debate, proponents of university interactions with industry contend, “Universities can and should play a larger and more direct role in assisting industry” without concern for fears of dilution or corruption of their research (Rosenberg and Nelson, 1994, p. 323).

Rosenberg and Nelson (1994) argue that universities should play a larger role developing solutions to the needs of industry contingent on a division of labor existing between university and industry. This division should allow universities to focus on problem solving but prevent them from making decisions based on commercial criteria. The authors explain, “In general, university researchers are poorly equipped for judging what is likely to be an acceptable solution to a problem and what is not [in commercial terms]” (p. 346). Even with a strictly maintained division of labor, university research may still be influenced in terms of the time horizon of projects. As explained by Rosenzweig (1999), if university researchers do focus on commercial problem solving, short-term research may begin to drive out long-term research of unpredictable value. Hence, Rosenberg and Nelson’s (1994) proposal may not pacify all the concerns of university-industry relationships critics.
The discussion of university-industry relationships is obviously not limited to economists. As reported by Dasgupta and David (1994) recently in Time Magazine, Walter Massey, then director of the National Science Foundation, is quoted as saying:

The public hears that we’re No. 1 in science, and they want to know why that fact isn’t making our lives better. The one thing that works in this country doesn’t seem to be paying off (p. 488).

Massey’s comment eludes to an important aspect of the university-industry relationship regarding knowledge transfer that economists have not investigated fully. Recent studies focus on overt knowledge transfer mechanisms; these include joint project collaborations, publications, patents, and conferences (Henderson, Jaffe, and Trajtenberg, 1998b, Agrawal and Henderson, 2002, Shane 2002). Less overt mechanisms of knowledge transfer that facilitate technical advance have been overlooked. These mechanisms have gone unobserved because surprisingly few studies examine the behavior of the university researcher as it relates to knowledge transfer. Many studies examine knowledge transfer by investigating conduits for transfer but neglect the motivations of the researchers creating the knowledge.

2.2 Incentives of Academic and Commercial Researchers

Graff, Heiman, and Zilberman (2002) explain that an academic researcher strives to maximize some combination of fame, fortune, and freedom. Similarly, an industry researcher also strives to maximize some combination of these three factors. The definitions of fame and fortune are obvious. Fame describes the prestige of a researcher while fortune describes the wealth of a researcher. The meaning of freedom, however, is not as apparent. Freedom relates to self-determination. It is the capacity for a researcher to define his own research objectives, strategies, and projects. As explained dramatically by Graff, Heiman, and Zilberman (2002), freedom is the desire of researchers “To be one’s own boss and let loose their passion for the
pursuit of scientific knowledge, enjoying ‘the love of the hunt’” (p. 92). Fame, fortune, and freedom are not mutually exclusive objectives. These three factors have some degree of dependence, so a change in one factor will likely influence another factor. The dependence among these factors becomes relevant when considering the incentives faced by university and commercial researchers.

The researcher working in a university environment and the researcher working in a commercial environment face distinct incentive structures (Audretsch and Stephan, 1999, Stern 1999). Stern (1999) explains these differing incentive structures are the result of two fundamental differences between academic and commercial research. First, academic research is focused on the formulation, testing, and solving of problems that do not need to have commercial application. Second, the rewards to solving problems in academia are primarily non-pecuniary, such publishing and gaining the admiration of colleagues. In contrast, the rewards in industry are associated with pecuniary rewards from monopolization of knowledge. These differences in research environments create distinct incentive structures in each environment which encourage researchers to maximize their utility differently via the factors of fame, fortune, and freedom.

Audretsch and Stephan (1999) elaborate on the differences between incentives structures in the university and commercial research environments. The academic environment rewards researchers for the creation and sharing of new scientific knowledge while the corporate environment rewards researchers for the creation of new economic knowledge but not necessarily scientific knowledge. By economic knowledge, Audretsch and Stephan (1999) refer to new information that has value in the marketplace. Consequently, the primary goal of the university researcher is to establish priority through publication while researchers in industry are discouraged from sharing new knowledge.
The differences in incentive systems in academic and commercial environments encourage a researcher to focus on developing research whose primary purpose after problem solving is to either increase the researcher’s fame or increase the researcher’s fortune. The freedom granted to a researcher is highly correlated with the other two factors within both the university and commercial settings. In the university environment, freedom follows closely with fame. University researchers who have a demonstrated prominence in their field of study are given more freedom in their research. In contrast, in the commercial environment, freedom follows closely with fortune. Commercial researchers who have produced research that has proven extremely profitable are given more freedom in their research projects. Because of freedom’s relationship to fame and fortune, a researcher will likely focus on maximizing fame or fortune depending on the environment and expect to receive increased freedom as a by-product.

Audretsch and Stephan (1999) find that the fundamental differences in academic and commercial incentive structures cause researchers in these institutions to have distinct career paths. Using life-cycle models, Audretsch and Stephan (1999) determine that university researchers invest heavily in developing their reputation earlier in their careers. Later in their careers, university researchers tend to cash in on their reputation investment for economic returns. In contrast, private sector researchers are more fully monetarily compensated for their production of knowledge throughout their careers. Consequently, maturity of career is a strong predictor of behavior for an academic researcher but not for an industry researcher.

2.3 Using Patents as Research Products

To understand how researcher behavior and the resulting research products change as a result of differing incentive structures, we ideally would want to evaluate the main products of research. For universities, this product is scholarship. Scholarship is frequently represented by
publications in academic journals (Agrawal and Henderson 2002, Weingart, Sehringer, Winterhager 1990). Critics of university-industry relationships fear that academic researchers face a different incentive structure for producing research when working with industry than when working in academia. These differing incentive structures may cause research to have a more pure or more commercial orientation.

Evaluating a publication on a scale of research purity and commercial orientation is a difficult task for a number of reasons. Foremost of these challenges is the difficulty in simply developing a procedure to systematically classify publications along these qualitative dimensions. Furthermore, once a system is constructed, obtaining the characteristic data for each publication presents another challenge. Consequently, it is necessary to find another research product that can appropriately capture the effects of different environmental and behavioral factors along the dimensions of purity and commercial orientation. Patent data provides the closest solution to the requirement of a quantitative evaluation of qualitative characteristics. Trajtenberg, Henderson, and Jaffe (1997) develop techniques that take advantage of patent citation information to quantitatively evaluate characteristics of patents. We employ these techniques to evaluate the patent data used in this study.

A patent is a document granted by an authorized government agency to an inventor that establishes a legal property right to an invention. In the United States, this granting agency is the United States Patent and Trademark Office (USPTO). The USPTO defines the right issued by a patent grant as “‘the right to exclude others from making, using, offering for sale, or selling,’ the invention in the United States or ‘importing’ the invention into the United States” (USPTO, 21 January 2003). Once a patent application has been submitted, it is evaluated by the USPTO according to two primary criteria: potential utility and novelty as claimed in the patent
application. Utility is defined as the usefulness of an invention while novelty is an invention’s uniqueness. Ultimately, a patent examiner decides if an invention fulfills these criteria.

A grant is issued to the inventor if the application sufficiently meets the utility and novelty criteria. The right to the invention, however, can either be assigned to the inventor or to another party, known as the assignee. As Griliches (1990) explains, a patent usually is assigned to the inventor’s employer or sold to another party. Once the patent has been granted, the right to the invention is protected by the threat and enforcement of infringement penalties. Through these processes, the patent system attempts to “Encourage invention and technical progress both by providing a temporary monopoly for the inventor and by forcing the early disclosure of the information necessary for the production of this item or the operation of the new process” Griliches (1990, p. 1663). Dam (1998) notes that from an academic researcher’s perspective the required invention disclosure is unlikely to contain significant scientific or technical data so is not as useful for continued research as journal publications. Patents, however, do capture the general concepts and solutions developed by an inventor.

2.4 Patent Process at Stanford

There is a caveat to using patent data as a proxy for university research products. As illustrated in Figure 1, all research at Stanford University is obviously not patented. Figure 1 categorizes research products at Stanford and depicts the grouping of the product subset of patents. From Figure 1, we see that patents are a small subset of all possible research products. Recognizing patents represent a subset of research products, it is important to understand how the patent process at universities filters research. Since all of the data used in this study relates to Stanford University, the process at Stanford is described. The patenting procedures at Stanford are similar to the protocols followed by most American universities.
At Stanford, the Office of Technology Licensing (OTL) administers the process of identifying and managing intellectual property assets. OTL policies encourage researchers to disclose an invention to the office through a legal and ethical obligation which they enter into through a signed document (Scholes, 6 December 2002). Of all possible research products, only a subset is commercially viable (See Figure 1). A commercially viable invention is an invention that the marketplace values. As illustrated in Figure 1, of these commercially viable inventions, some may not be disclosed to the OTL. One of the reasons a Stanford researcher would not disclose to the OTL is it may be more profitable to disclose outside of Stanford through a private company.

For those researchers who choose to disclose to Stanford, the process begins when the researcher submits an “Invention and Technology Disclosure Form” to the OTL. If senior staff at the OTL perceives the invention as having commercial viability, a docket is created. A docket can be understood as an organizational mechanism that contains all information related to the
invention including: descriptive information, the inventors involved, the sponsors of the work, public disclosures, and so forth.

Once a docket has been created it is assigned to an associate at the OTL who is responsible for handling all actions relating to the docket. The associate meets with the inventors to jointly discuss the invention. Through these discussions the associate evaluates the manufacturing feasibility, novelty, potential applications, and possible markets for the invention. Based on these meetings, the associate develops a preliminary licensing strategy.

An important aspect of the licensing strategy determined by the associate is if Stanford will elect to take title of the invention and file for a patent on the invention. Because of the high cost of filing patent applications, the OTL prefers to have potential licensees before committing to protection in the form of patent. If a patent application is filed and granted, the OTL then licenses the invention to companies that the associate and inventor identify as having relevant interest in the invention.

After a licensing agreement is reached between the OTL and a company, the royalty payments are collected by the OTL. After expenses such as legal fees are deducted from the payments, 15% of the royalties are distributed to the OTL. The remainder of the royalties is divided three ways: one-third to the inventor, one-third to the inventor’s department, and one-third to the inventor’s school. (OTL, 21 January 2003; Scholes, 6 December 2002).

Consequently, by fulfilling their obligation to disclose to Stanford, researchers are likely to receive lower monetary gains than if they had disclosed outside of Stanford.

2.5 Measuring Patent Characteristics

Two characteristics of patents can be measured from citations data: *basicness* and *appropriability*. The methodology of characterizing patents along these dimensions through
citation information was first developed by Trajtenberg, Henderson, and Jaffe (1997). The authors originally developed the measures to examine characteristics of inventions from different research environments. Trajtenberg, Henderson, and Jaffe (1997) chose to categorize patents according to basicness and appropriability because both are major sources of heterogeneity in the economics of technical change. Basicness relates to the purity of an invention; the term “Refers to fundamental features of innovations such as originality, closeness to science, breadth, etc, that impinge on the incentives to engage in R&D and on the choice of research projects” (Trajtenberg, Henderson, and Jaffe 1997, p. 51). Appropriability “Refers to the ability of inventors to reap the benefits from their own innovations” (Trajtenberg, Henderson, and Jaffe 1997, p. 51).

Along the dimensions of basicness and appropriability, Trajtenberg, Henderson, and Jaffe (1997) compare the patents assigned to universities to the patents assigned to industry. From their comparison, they generate measures of basicness and appropriability. The fundamental assumption that they employ to develop measures of basicness and appropriability is that research produced from universities is distinctly different from research produced in industry. Universities are thought to produce more basic, less appropriable research, and industry is thought to produce less basic, more appropriable research. From this assumption, the authors create three measures: Generality, Importance, and Self-Citations. To understand the value of these measures, it is important to understand the innovation measurement problem as described by Kuznets (1962):

Some inventions, representing as they do a breakthrough in a major field, have a wide technical potential in the sense that they provide a base for numerous subsequent technical changes. In this respect the invention of the first steam engine, which initiated a whole series of major technical changes and applications…subsidiary inventions…is vastly different from the invention of the safety match….This wide range is for our purposes the major characteristic
relevant to the problem of measurement (p. 26).

Through the use of patent citations, the measurement problem can be resolved. A systematic, quantitative distinction can be made between a steam engine and a safety match.

The basicness of an invention is estimated through the relationship between a patent and its descendents. A patent’s descendents are patents that cite the original patent. The two instruments that measure basicness are Generality and Importance. Generality gauges the degree to which an invention’s subsequent technological advances are spread across different fields.

\[
\text{Generality}_i = 1 - \sum_{k=1}^{N_i} \left( \frac{N_{\text{citing}}_{ik}}{N_{\text{citing}}_i} \right)^2
\]

where \( N_{\text{citing}} \) is number of patents citing the originating patent

\( i \) is the patent

\( k \) is the index of patent classes

\( N_i \) is the number of different classes to which the citing patents belong

Therefore, the larger Generality, the less technologically concentrated are the patent’s descendents.

The second measure of basicness, Importance, estimates the technological significance of an invention as a function of patents citing it and these patents’ respective importance.

\[
\text{Importance}_i = N_{\text{citing}}_i + \lambda \sum_{j=1}^{N_{\text{citing}}_i} N_{\text{citing}}_j
\]

where \( N_{\text{citing}} \) is number of patents citing the originating patent

\( i \) is the patent

\( j \) is the citing patent

\( \lambda \) is a discount factor that is meant to down-weight the second-generation descendents relative to the first-generation descendents

Henderson, Jaffe, Trajtenberg (1998b) and Mowery and Ziedonis (2002) exclude the discount factor, \( \lambda \), because of the brief period of time following the issuance of most university patents.
The authors argue that the discount factor has no effect on the importance score calculation. The time distribution of the data in this study is discussed in the Data section. Consequently, \textit{Importance} is defined as simply the number of citations received by the original patent. Hence, the more citations a patent receives, the more important the patent.

The second dimension along which an invention can be evaluated is appropriability. Appropriability of an invention can be measured through an instrument that relies on the ownership structure of an invention as evinced by patent citations. As Trajtenberg, Henderson, and Jaffe (1997) explain, \textit{Self-citation}, is defined “As the percentage of citing patents issued to the same assignee as that of the originating patent” (p. 62). The basis for this instrument is that citing patents will reflect subsequent developments to the original invention. These developments are the channel through which returns are appropriated to the assignee.

\[ \text{Self-Citation} = \frac{N_{citing_a}}{N_{citing_i}} \]

where \( N_{citing} \) is number of patents citing the originating patent
\( i \) is the patent
\( j \) is assignee

In addition to a citation approach to determining appropriability, a second measure can also be employed that is correlated to appropriability. The licensing revenue of invention provides another measure of the assignee’s ability to reap the profits generated by the invention. As the licensing revenue of an invention increases, so does its level of appropriability.

\textbf{2.6 Measures as a Proxy for Economic and Technology Concepts}

Using Trajtenberg, Henderson, and Jaffe’s (1997) citation measures and licensing revenue data, it is possible to understand the characteristics of inventions, which are a subset of academic research products. To apply the citation measures within a university context, it is necessary to evaluate the assumptions that Trajtenberg, Henderson, and Jaffe (1997) make to
construct the measures. The authors’ main assumption is that university research is distinctly different from industry research along the dimensions of basicness and appropriability. This assumption seems to have validity in light of the arguments of Rosenberg and Nelson (1994). Rosenberg and Nelson (1994) explain that research at American universities has a long-term time orientation and aims to understand phenomena at a fundamental level, while research in industry has a short-term time orientation and is focused on understanding problems relevant to the marketplace. Because of the validity of this assumption and previous economic literature that employs these measures in a similar manner (Henderson, Jaffe, Trajtenberg 1998b, Mowery et al. 2001, Mowery and Ziedonis 2002), the patent citations measures developed by Trajtenberg, Henderson, and Jaffe (1997) can be employed to characterize academic patents.

Before the patent citation measures can be employed, the interpretation of the measures along the dimensions of basicness and appropriability within the context of university research should be understood. Basicness refers to an invention’s position on a scale that ranges from more pure to more applied. University research can span this range of basicness as it is defined in Rosenberg and Nelson (1994). Although many applied inventions do indeed have high commercial returns, it is also possible for a pure invention to have high commercial returns. For example, recombinant DNA cloning is considered a basic scientific invention but has generated tremendous returns for its inventors and Stanford University. The commercial value of an invention is best measured by appropriability. Hence, recombinant DNA cloning research is basic and appropriable.

The more basic an invention, however, the less likely the invention will have a patent. Maxwell’s equations are an example of basic research that is not patentable. Consequently, as illustrated earlier in Figure 1, the most fundamental research is excluded by using patent data.
Within the basicness dimension, although the measures of *Generality* and *Importance* are related they describe two different characteristics of a patent. For example, some inventions may have a low *Generality* score, meaning they are not frequently cited by patents from many different technological categories, but represent basic research. Consequently, another measure of basicness is necessary to appropriately classify inventions from the low *Generality* score but basic case. This is the rationale for employing *Importance* in conjunction with *Generality* to proxy basicness.

As *Importance* increases, Trajtenberg, Henderson, and Jaffe (1997) argue the basicness of an invention also increases. For example, the ruby laser is a highly cited patent that does not have direct applications, meaning it is pure in its fundamental form. When coupled with other inventions, the ruby laser becomes less basic and more readily applied to problems. The act of being coupled to other inventions is reflected in the number of cites the ruby laser received—its *Importance*. The *Generality* of the laser is also captured through citations. Since the laser is cited across many different technological categories, its use is beyond a directly related research area. This quality is an aspect of more pure forms of research whereas less pure research tends to focus on applications that are limited to directly related fields. Figure 2 illustrates the overlapping nature of these characteristics as they relate to the purity of research.
As research becomes more general and important it becomes more pure. From Figure 2, research that receives a high *Importance* score but low *Generality* score is not considered pure research. Research with these characteristics is cited a great deal only within its specific technological category such as a process developed by Stanford researchers for recovering nucleated fetal cells. Although the research is high cited, it is only relevant within one specific area of research.

Although the distinction is subtle, pure research should not be defined as university research. Pure research is a subset of university research since the end goal of academic research is to understand a “phenomena at a fundamental level” (Rosenberg and Nelson, 1994). By this definition, some university research may not be pure at all but instead narrowly applied. Regardless of the level of basicness of university research, it should have a high *Importance* score. Since the research is describing a process at a fundamental level, it may be more applied in its orientation but should still have a broad enough scope to receive a high number of citations, as in the nucleated fetal cell case.
Independent from the basicness of research is the appropriability of research. Appropriability of research can be high for basic as well as for applied research. University research should generally have less appropriability than industry research. University research is orientated towards solving problems for the sake of solving problems whereas industry research solves problems deemed valuable by the market place. Both of the measures of appropriability, Self-Citation and Licensing Revenues, capture the same type of information. But, the lack of licensing information on every patent, makes Self-Citation a more easily obtained measure.

3 MODEL

3.1 Research Questions

To understand the behavior of university professors, we explore two interrelated questions. First, we explore the relationship between the researcher’s environment and characteristics of his research. Second, we examine the career anticipatory behavior of a researcher. Each of these topics is investigated using patent data as a proxy for research and patent citation measures as well as limited licensing revenues data as proxies for characteristics of the research.

3.2 Utility Model

Considering the research completed by Graff, Heiman, and Zilberman (2002) and Audretsch and Stephan (1999), a simple, formal utility model of the behavior of a researcher in either a university or commercial environment can be constructed. The prestige level of a researcher is directly related to the number of citations his published writings receive. The more important the research, the more academic citations a paper receives, the higher the prestige level of the researcher. The commercial value of a research project is related to the appropriability of the research. For convenience, let \( I \) represent the importance of a research project and let \( A \)
represent the appropriability of a project. In addition to characteristics of the research project, the wealth of the researcher, \( Y \), and the capabilities of the researcher, \( \alpha \), influence a researcher’s utility. The capabilities of a researcher, \( \alpha \), refer to the heterogeneous capabilities and skills of a researcher that influences his preferences to produce certain types of research. These factors can be taken as inputs to a researcher’s utility function:

\[
U(I, A, Y, \alpha)
\]  \hfill (1)

Where

- \( U \) is the researcher’s utility
- \( I \) is the importance of the research project
- \( A \) is the appropriability of the research project
- \( Y \) is the wealth of the researcher
- \( \alpha \) is the capabilities of the researcher

A researcher will maximize his utility subject to some budget constraint represented by his wealth, \( Y \). The budget constraint faced by a researcher is a function of the incentives of his research environment. These incentives change depending on the research environment because of variations in reward systems. The incentives are not likely to change because of differences between equipment and facilities in academic and commercial environments. The equipment and facilities that are accessible by a leading academic researcher are also accessible by a leading commercial researcher. Consequently, the budget constraint can be written in the following functional form:

\[
Y(I, A, \theta)
\]  \hfill (2)

Where

- \( Y \) is the wealth of the researcher
- \( I \) is the importance of the research project
- \( A \) is the appropriability of the research project
- \( \theta \) is the institution or funding source supporting the researcher
Considering the utility function of a researcher, it is possible to determine how he maximizes his utility function, equation (1), subject to his budget constraint, equation (2):

\[ \max U(I, A, Y, \alpha) \text{ subject to } Y(I, A, \theta) \]  

(3)

Through equation (3), we are able to solve for \( I \) and \( A \), where:

\[ I(\alpha, \theta) \]  

(4)

\[ A(\alpha, \theta) \]  

(5)

From equations (4) and (5), we can see that the importance, \( I \), and the appropriability, \( A \), of a research project are functions of the researcher’s capabilities and the supporting institution. This implies that as the research environment varies between academic and commercial, the research produced will respond to the differences in the environment. A researcher will maximize his utility by producing research that is most valued according to the research environment, \( \theta \), and his unique capabilities, \( \alpha \).

### 3.3 Effect of Environment on Research Characteristics

Because of the differences in incentives faced by researchers in academic and commercial environments, research produced in these environments should be distinct. The incentive structure of the university does not reward wealth maximizing research (Audretsch and Stephan, 1999). A researcher in the university environment is encouraged to produce research that results in a high return on fame. In contrast, the incentive structure of industry does not reward fame maximizing research. Consequently, the same researcher in a commercial environment is encouraged to produce research that results in a high return on wealth.

To determine how the behavior of a researcher and thereby her research responds to changes in environment, we test two hypotheses. First, we test if research produced in an academic environment is different from research produced in a commercial environment. To
measure differences in research, we observed two characteristics of research: importance, $I$, and appropriability, $A$. As discussed earlier, importance refers to the number of citations received by a published work. As these citations increase, the prestige of the researcher also increases. Appropriability refers to the value of a research project as evaluated by the marketplace. If the behavior of a researcher is independent of institution, the degree of importance and the level of appropriability of the research should be consistent regardless of whether the researcher is working in a university or commercial lab:

$$I_{ji}(a_i, \theta_A) = I_{ji}(a_i, \theta_C) \quad (6)$$

and

$$A_{ji}(a_i, \theta_A) = A_{ji}(a_i, \theta_C) \quad (7)$$

Where

$I_{ji}$ is the importance of project $j$ for researcher $i$

$A_{ji}$ is the appropriability of project $j$ for researcher $i$

$a_i$ is the preferences of researcher $i$

$\theta_A$ is an academic institution

$\theta_C$ is a commercial institution

If the behavior of a researcher is dependent on institution, the research produced in each environment should be different. In particular, considering the incentives structures of academic and commercial institutions, we would expect the following:

$$I_{ji}(a_i, \theta_A) > I_{ji}(a_i, \theta_C) \quad (8)$$

and

$$A_{ji}(a_i, \theta_A) < A_{ji}(a_i, \theta_C) \quad (9)$$

Where

$I_{ji}$ is the importance of project $j$ for researcher $i$

$A_{ji}$ is the appropriability of project $j$ for researcher $i$

$a_i$ is the preferences of researcher $i$

$\theta_A$ is an academic institution

$\theta_C$ is a commercial institution
Consequently, we hypothesize that research completed by the same researcher in an academic environment is more important and less appropriable than research completed in a commercial environment.

The second hypothesis we test investigates if different funding sources affect the behavior of an academic researcher. Within a university, there are three primary sources of funding for research: internal university funding, federal funding, and corporate funding (Hall, 2002). These distinct funding sources often have different objectives for supporting academic research. For example, research supported by the National Science Foundation may have a broad scope whereas research supported by the SHARP Corporation is focused on solving a specific problem.

These differences in objectives may affect the incentives faced by academic researchers. If the incentives faced by an academic researcher change depending on the source of funding, then the characteristics of the research will change. The two characteristics of research measured are again importance and appropriability. If sponsorship does not influence the incentives of a researcher this implies that:

\[ I_{j(i)}(\alpha_i, \theta_A) = I_{j(i)}(\alpha_i, \theta_G) = I_{j(i)}(\alpha_i, \theta_C) \]  
(10)

and

\[ A_{j(i)}(\alpha_i, \theta_A) = A_{j(i)}(\alpha_i, \theta_G) = A_{j(i)}(\alpha_i, \theta_C) \]  
(11)

Where

- \( I_{j(i)} \) is the importance of project \( j \) for researcher \( i \)
- \( A_{j(i)} \) is the appropriability of project \( j \) for researcher \( i \)
- \( \alpha_i \) is the preferences of researcher \( i \)
- \( \theta_A \) is academic funding
- \( \theta_G \) is government funding
- \( \theta_C \) is commercial funding

If sponsorship does influence the incentives of a researcher, it is likely that the reward systems of each sponsoring organization will lead to different levels of importance and appropriability.
Considering the earlier discussion of reward systems, research may differ according to sponsorship as follows:

\[ I_j(i, \theta_A) > I_j(i, \theta_G) > I_j(i, \theta_C) \]  
\[ A_j(i, \theta_A) < A_j(i, \theta_G) < A_j(i, \theta_C) \]

Where

- \( I_j(i) \) is the importance of project \( j \) for researcher \( i \)
- \( A_j(i) \) is the appropriability of project \( j \) for researcher \( i \)
- \( \alpha_i \) is the preferences of researcher \( i \)
- \( \theta_A \) is academic funding
- \( \theta_G \) is government funding
- \( \theta_C \) is commercial funding

Consequently, an academic researcher may favor producing research that has a high degree of appropriability for commercially supported projects and may favor producing highly cited research if sponsored only by university funds. We hypothesize that different sources of funding will influence a researcher’s utility function and alter the characteristics of research conducted within a university.

3.4 Career Anticipatory Behavior of Researcher

The second research question investigates the career anticipatory behavior of a professor. As explained earlier, there are three factors that determine the behavior of professors: fame, fortune, and freedom. Some professors, as found by Audretsch and Stephan (1999), plan to leave their work in academia and become entrepreneurs. For these professors, maximizing their level of fortune enables them to reach the highest level of utility. Audretsch and Stephan (1999) using life-cycle models find that professors invest heavily in increasing their fame early in their careers to later cash in for fortune later in their careers. The entrepreneurial professor is someone who conducts academic research but then later also conducts commercial research building on her academic expertise. These entrepreneurial professors work in an academic
environment but can have a large increase in their future utility from producing appropriable research, just as a commercial researcher. If a professor’s objective is to maximize wealth, the most efficient path to accomplishing this objective is to produce research that is deemed valuable by the marketplace. Consequently, the entrepreneurial professor while producing fame building research will also produce precursory tool building research whereby they develop physical products, systems, and expertise valued by the marketplace. By creating these tools, professors develop assets that have value to them as entrepreneurs. We investigate the degree to which this tool building research occurs to determine if professors anticipate their future career roles.

We can determine the existence of the tool building by examining the relationship between precursory academic research and later commercial research. If tool building is observed, there are two likely behavioral interpretations. First, a professor early in his career has the intention to become an entrepreneur. As a result, the professor responds to the incentive structure of his future environment, \( \theta \); he develops tools that will later facilitate entrepreneurial activities. Second, a professor unintentionally develops tools that are commercially valuable. The development of these tools then leads the professor to pursue entrepreneurial activities. The professor’s preferences given a set of capabilities, \( \alpha \), change. The professor produces different types of research as \( \alpha \) changes. In both cases, however, a professor is aware of the commercial value of the tools and pursues entrepreneurial activities using these tools.

We can use the utility function in equation (1) to understand the anticipatory behavior of the entrepreneurial professor that is associated with tool building. If an area of a professor’s academic research investigates topics which have a high degree of market value, the professor will want to maximize her pecuniary returns to her research. To maximize her returns, the professor will take the tools developed on a specific research area in the academic environment
and capitalize on them in a commercial environment. This professor is maximizing her utility over her lifetime by responding to incentives of an anticipated commercial work environment, $\theta_E$. Because the entrepreneurial professor is aware of the research’s value by the marketplace and its the potential pecuniary returns, her utility is higher than the non-entrepreneurial professor working in the same market valued area. As a result of the differences in utility, research in the same field for the entrepreneurial professor and non-entrepreneurial professor will have the following relationship:

$$A_{j(E)}(a_E, \theta_E) \geq A_{j(NE)}(a_{NE}, \theta_{NE})$$

(14)

Where

- $A_{j(E)}$ is the appropriability of project $j$ for the entrepreneurial researcher
- $A_{j(NE)}$ is the appropriability of project $j$ for the non-entrepreneurial researcher
- $a_E$ is the capabilities for the entrepreneurial researcher
- $a_{NE}$ is the capabilities for the non-entrepreneurial researcher
- $\theta_E$ is the anticipated work environment of the entrepreneurial researcher
- $\theta_{NE}$ is anticipated work environment of the non-entrepreneurial researcher

A subtle but critical aspect of the anticipatory behavior hypothesis addresses the variation in research activities of entrepreneurial professors. Investigating the existence of tool building does not examine if a professor is failing to fulfill his academic research obligations by instead completing commercially valuable research. Within the data, it is not possible to determine how a professor is dividing his time in terms of producing commercially useful and academically useful research. Considering the capabilities of the model, however, we predict that tool building does occur among entrepreneurial professors.

4 DATA

The behavioral model developed in this paper relates the characteristics of a research project to the actions of a researcher as they are influenced by the researcher’s environment and anticipatory behavior. To assess this model, two sets of data were collected. The first set of data
was compiled by Hall, Jaffe, and Trajtenberg (2001) through research sponsored by the National Bureau of Economic Research (NBER). The second data set was acquired through the Office of Technology Licensing at Stanford University.

4.1 NBER U.S. Patent Citations Data File

The NBER U.S. Patent Citations Data File data set represents an enormous data collection endeavor that was recently completed and made publicly available. The dataset consists of 3 million utility patents that range from January 1, 1963 to December 20, 1999. Although there are three main categories of patent documentation—utility, design, plant—the overwhelming majority of patents are classified as utility patents. As explained earlier, a utility patent is used to protect an invention that is novel due to function or process.

The main strength of the NBER dataset is that it facilitates searching for patents by citations. It is possible to determine all citations to a patent before 1999 without a great deal of effort. The second strength of the NBER dataset is the volume of original information it includes for each patent, which is presented in Hall, Jaffe, and Trajtenberg (2001). The key weakness of the NBER dataset is that it has only modest information about the inventors listed on the patent. The dataset also lacks funding information and appropriability scores for each patent.

4.1.1 Original Patent Variables in NBER Dataset

Relevant information contained in the NBER dataset includes:

*patent* = patent number, a unique numerical identifier for each patent filed in the United States. The patent number is used as a key variable to merge the two sets of data.

  Example: 5,510,260

*gyear* = grant year, the year in which the patent was granted.

  Example: 1996
appyear = application year, the year in which a patent was applied for. The application year is
closer than gyear to the timing of when an invention was actually created due to the substantial
processing times at the Patent Office.
    Example: 1993

nclass = technological category, a three-digit classification code developed by the USPTO that
groups inventions into technology groups.
    Example: 435 = Tobacco extracts made by fermentation and manufacturing
    processes

lastname = inventor last name
    Example: Zare

firstname = inventor first name
    Example: Richard

midname = inventor middle name or initial
    Example: S

4.1.2 Constructed Variables in NBER Dataset

Hall, Jaffe, Trajtenberg (2001) construct a number of variables from the original patent
information that are useful in evaluating a patent. These constructed variables include:

cat = primary technological category, a one-digit classification code developed by Hall, Jaffe,
    Trajtenberg (2001) that provides a more general classification scheme than nclass.
    Example: 3 = Drugs and Medical

subcat = technological subcategory, a two-digit classification code developed by Hall, Jaffe,
    Trajtenberg (2001) that provides a more general classification scheme that the USPTO but more
    specific than cat.
    Example: 33 = Biotechnology

general = measure of generality, a estimate of the degree to which subsequent technical advances
to an invention are spread across different technological fields.
    Example: .245

importnc = measure of importance, an estimate of the technological significance of an invention
that counts citations received by a patent. The more citations received, the more important the
patent.
    Example: 14
4.2 OTL Data File

The second dataset, which was collected through the OTL, contains information on patents that have been assigned to Stanford. This data is part of a proprietary database managed by the OTL. The OTL utilizes this database to administer its caseloads and document all interactions between Stanford and non-Stanford clients. The strength of the OTL dataset is that it cross-links information through the docket organizational structure. The docket structure links funding, department, and other information to patent numbers. The main weakness of the dataset is the lack of licensing revenue data for Stanford patents.

4.2.1 Original Variables in OTL Dataset

This dataset consists of the following variables:

- **docknum** = docket number, a docket is an organizational structure used by the OTL. Each docket contains a variety of information on an invention and links this information together. Note, every docket does not contain a patent but each docket does relate to one invention with perceived commercial viability.
  
  Example: 85-109

- **patent** = patent number, a unique numerical identifier for each patent filed in the United States. The patent number is used as a key variable to merge the two sets of data.
  
  Example: 5,510,260

- **sponsor** = name of organizations that were disclosed as sponsors. Every invention disclosed to the OTL is required to list any sponsorship of the research that led to the invention. For inventions that had no listed sponsorship, Stanford was the only source of funding.
  
  Example: National Institutes of Health

- **sppontype** = classification index for sponsors. Every sponsorship type is classified into numerical category corresponding to a sponsorship type such as private, government, non-profit, only Stanford, and mixed. Government is broken down into subcategories such as NIH and military.
  
  Example: Stanford = 1

- **revenue** = total licensing revenue over the life of a patent to December 1999. The OTL strictly monitors licensing revenues for all of the inventions disclose to it by Stanford inventors.
  
  Example: $12,000,000
school = name of school that corresponds to each listed inventor.
    Example: School of Engineering

schtype = classification index for school.
    Example: School of Engineering = 2

lastname = inventor last name
    Example: Zare

firstname = inventor first name
    Example: Richard

midname = inventor middle name or initial
    Example: S

4.3 Merged Dataset

By merging the NBER and OTL datasets a great deal of information is linked. In particular the patent portfolios for every Stanford researcher who has ever disclosed to Stanford can be constructed. This aspect of the data provides us with the opportunity to carry out a unique investigation. The patent portfolios for researchers allow us to conduct within-subject experimental design procedures. By examining the research carried out by the same individual inside and outside of Stanford, we are able to embrace the unobserved heterogeneity among individual subjects that is often a burden in such studies. In addition to this advantage, the funding support that led to a particular patent is link to characteristics of that patent. The main weakness of this dataset is that it is limited to only those Stanford researchers who have disclosed at least once to the OTL. This selection bias, however, is not likely to have a major impact on any of the hypotheses we investigate due to its direction. We exclude researchers that are “true entrepreneurs” so the results of our study would only be strengthened if we observed these researchers. A second weakness in the merged data is the size of the patent portfolios of Stanford researchers. Although some researchers have well over 100 patents, others have less
than 5. Because of the small sample size for each researcher, we cannot assume normality. This weakness can be easily resolved through using nonparametric analysis procedures.

4.3.1 Constructed Variables in Merged Dataset

In addition to the variables in the NBER and OTL datasets, we construct a variable using techniques developed in Trajtenberg, Henderson, and Jaffe (1997) that takes advantage of the merged data:

\[ selfcite = \text{measure of appropriability, an estimate of the market returns to an invention that examines the ownership structure of a series of citing patents.} \]

Example: .15

4.4 Data Characteristics

Before a formal empirical analysis of the model is completed, we examine the characteristics of the data sample. Table 1 presents basic descriptive statistics about Stanford and non-Stanford patents. A Stanford patent is a patent where a Stanford researcher is listed as the inventor and is disclosed to Stanford. A non-Stanford patent is also a patent where a Stanford researcher is listed as the inventor, but it is disclosed to an organization other than Stanford. This organization is usually a private company.
Table 1: Descriptive Statistics for Stanford and Non-Stanford Patents

<table>
<thead>
<tr>
<th></th>
<th>Stanford</th>
<th>Non-Stanford</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Patents</strong></td>
<td>867</td>
<td>5183</td>
</tr>
<tr>
<td><strong>Generality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.33</td>
<td>.25</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.29</td>
<td>.29</td>
</tr>
<tr>
<td>Median</td>
<td>.33</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>.84</td>
<td>.91</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Importance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.2</td>
<td>6.0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>9.25</td>
<td>11.65</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Maximum</td>
<td>138</td>
<td>292</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Appropriability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.09</td>
<td>.09</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.22</td>
<td>.21</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Generality and Appropriability scores are constrained to a range of 0 to 1 while Importance scores only have a lower bound constraint of 0.

From Table 1, the skewed distributions of the merged data become apparent. A general trend in the citations measures is a clustering near zero. This clustering is clearly illustrated in the distribution plots for each of the citation measures (Figure 3-6). The implications of the clustering around zero are different for each measurement.
4.4.1 Generality Score Distribution

**Figure 3a: Distribution of Generality Scores for Stanford Patents**

```
Grouped Generality Score

.8-.9 120
.7-.8 100
.6-.7 80
.5-.6 60
.4-.5 40
.3-.4 20
.2-.3 0
.1-.2 0

Patent Count
```

**Figure 3b: Distribution of Generality Scores for Non-Stanford Patents**

```
Grouped Generality Score

.9 700
.8 600
.7 500
.6 400
.5 300
.4 200
.3 100
.2 0
.1 0

Patent Count
```

Note: In the above plots (Figure 3a and 3b), all cases that received a zero score have been removed.
In Figures 3a and 3b, a generality score will equal 0 if a patent has received 0 or 1 citation from a future patent. In Figure 3a for Stanford patents, of the 488 patents whose generality score equals 0, 258 received 0 citations and 230 received 1 citations. In Figure 3b for non-Stanford patents, of the 3220 patents whose generality score equals 0, 1887 received 0 citations and 1333 received 1 citation.

From interpreting the definition of generality, the large number of Stanford and non-Stanford patents that received low generality scores means many of the patents have not been frequently cited from a number of different technological categories. One implication of this commonality is that some researchers in both Stanford and non-Stanford settings produce focused research that is not applicable to areas of study outside their own. The generality score of a patent can vary from 0-1 in Stanford and non-Stanford settings since a researcher in both environments does not face any systematic incentives to produce more general or less general research.
4.4.2 Importance Score Distribution

Figure 4a: Distribution of Importance Scores for Stanford Patents

Figure 4b: Distribution of Importance Scores for Non-Stanford Patents

Note: In Figure 4b all cases that received a zero score have been removed.
In Figures 4a and 4b, the clustering around 0 indicates that a number of patents owned by Stanford have not received any citations from other patents. From Table 1, the mean importance score of a Stanford patent is surprisingly lower than the mean importance score of all non-Stanford patents. We would expect that the average importance score for Stanford patents would be higher than for non-Stanford patents because of the differing incentives embedded in the utility functions for academic and commercial researchers.

### 4.4.3 Self-Citation Score Distribution

**Figure 5a: Distribution of Self-Citation Scores for Stanford Patents**

![Graph showing distribution of self-citation scores for Stanford patents.](image-url)
Note: In the above plots (Figure 5a and 5b), all cases that received a zero score have been removed.

In Figure 5a and 5b, a zero score is the result of two possible causes. First, if a patent never receives a citation, then the self-citation score is 0. Second, if none of the citing patents share the same assignee as the original patent, then the self-citation score will also be zero. In addition to the clustering, the large number and variation of co-assignees for patents owned by Stanford was unexpected. Of the Stanford patents, there are 441 unique co-assignees that were organizations, such as the SHARP Corporation and the US Navy. Distinct from this group, 105 other patents were jointly assigned to the individual inventors and to Stanford. The variations in the distributions of Stanford and non-Stanford patents indicate that research may differ in terms of appropriability depending on the work environment.
4.4.4 Licensing Revenues Distribution

In addition to using patent citations to estimate the appropriability of a patent, the licensing revenues for some Stanford patents are employed as a second measure of commercial viability of research. In Figure 6, the total licensing revenues from 1975-1999 for 521 of the 880 patents are presented. Unfortunately, due to difficulties in accessing bookkeeping records at the OTL, licensing data for each Stanford patent cannot be acquired. Of the 521 patents, only 100 have revenues totaling more than $200,000. Some of the patents with revenues greater than $200,000 had revenues well above this level; of these high earning patents, some had revenues in the millions. In 2000, the OTL claimed to have generated a cumulative income of approximately $450 million from all of its licensing activities, which includes patents.
4.4.5 Time Distribution of Patents

In addition to being aware of the overall distribution of the patent citation scores of Stanford researchers, it is important to consider the distribution of these researchers’ patents over time.

Table 7a: Time Distribution of Stanford Patents
Table 7b: Time Distribution of Non-Stanford Patents

![Graph showing the time distribution of non-Stanford patents from 1975 to 1999.](image)

From the Figures 7a and 7b, the skewed time distributions of Stanford and non-Stanford patents is apparent. After 1980, the growth rate of patents in both groups increase rapidly. A reason that at least partially explains the increase in patenting after 1980 is the Bayh-Dole Act. As explained by Mowery and Ziedonis (2002), the Bayh-Dole Act enabled universities to retain title to inventions which were developed using federal funding. The intent of Bayh-Dole was to encourage the adoption and use of inventions developed from federal support. This deregulation partially led to an increase in patenting among universities.

Other factors which Mowery and Ziedonis (2002) cite as contributing to the increase in patenting at Stanford specifically are the growing fields of biotechnology at universities and the Diamond versus Chakrabarty Supreme Court Decision which set an important precedent. The biotechnology field became an important and productive area of research at Stanford during the
mid 1970s and began to bloom over the next two decades. The Diamond versus Chakrabarty decision set a precedent which provided for the patenting of organisms, molecules, and research techniques emerging from biotechnology research. These factors combined with the general trend of an increase in patenting explain the relatively enormous number of recent patents.

When examining the distribution of the data, it is important to consider a truncation bias that exists in patent citations and also licensing revenues. The bias is the result of later patents not existing as long as earlier patents. All the citations measures are based on future citations. Consequently, there is a bias against more recent patents. For example, a patent issued in 1975 has existed for 10 more years than a patent issued in 1985. Simply because of time, the 1975 patent is more likely to have more citations to it than the 1985 patent. Similarly, licensing revenues for a patent tend to increase over time, so the 1975 patent will likely have greater total revenue than the 1985 patent. Since the Stanford and non-Stanford patents are sampled from the same time frames, both samples exhibit the same degree of truncation bias.

5 ANALYSIS

To test each of the three hypothesis developed in the Model section, we use categorical data analysis methods. We first investigate the effect of environment on research characteristics through the institution and sponsorship hypothesis. We then study the anticipatory career behavior of a researcher. For each hypothesis, we included a three part discussion which 1) expands on the theoretical foundations leading to a statistical test, 2) explains the results of the test given our statistical assumptions hold, and 3) investigates the implications of the testing if these assumptions are relaxed.
5.1.1 Effect of Environment on Research Characteristics, Institution Hypothesis

To determine if research characteristics change as a function of work environment, we expand on the theoretical behavioral model of a researcher to determine the appropriate method of statistical inquiry. From our behavioral theory, importance and appropriability research characteristics are a function of a researcher’s capabilities, $\alpha$, and the work environment of the researcher, $\theta$. In addition to these factors, the characteristics of a given research project may also be influenced by the moment in history when the research is taking place, $T$, and the research area of project, $D$. The importance and appropriability levels of a research project are determined relative to other research projects.

Considering these factors, a statistical model can be constructed to determine if research characteristics vary with environment. Employing categorical data analysis techniques, the model for each characteristic appears as:

\[
I_{j(i)} = B(\alpha_i + \theta_{E(i)} + T_j + D_j)
\]  
(15)

and

\[
A_{j(i)} = \Gamma(\alpha_i + \theta_{E(i)} + T_j + D_j)
\]  
(16)

Where

- $i$ is a researcher
- $j$ is a project of researcher $i$
- $\alpha_i$ is the capabilities of researcher $i$
- $\theta_{E(i)}$ is the work environment of researcher $i$
- $T_j$ is the time period when project $j$ is researched
- $D_j$ is the field of research of project $j$
- $I_{j(i)}$ is the importance of a project $j$ for researcher $i$
- $A_{j(i)}$ is the appropriability of a project $j$ for researcher $i$

From this model, we determine if $I$ and $A$ for a project vary with the work environment of the researcher, $\theta_{E(i)}$. 
To investigate this hypothesis, a patent portfolio for Stanford researchers is used. For each Stanford researcher the characteristics of her Stanford and non-Stanford patents is determined using patent citations measures. These characteristics are quantified estimates of the patent’s importance, $I$, and appropriability, $A$. As illustrated by the subscript $E(i)$ on $\theta$, Stanford patents represent one type of research environment and non-Stanford patents represent another. For the majority of Stanford researchers, the patents not pursued at Stanford were pursued in commercial settings. Hence, our sample consists of four groups of observations for each researcher: Stanford patents’ importance scores, non-Stanford patents’ importance scores, Stanford patents’ appropriability scores, and non-Stanford patents’ appropriability scores. If no significant difference exists between importance scores or appropriability scores for Stanford and non-Stanford patents, this implies that the behavior of the researcher is independent of the work environment. If a difference exists between importance scores or appropriability scores for Stanford and non-Stanford patents, however, this implies that the work environment does have an effect on the behavior of the researcher and thereby the characteristics of his research.

As discussed in the Data section the small sample size of patents for each researcher prevents us from making an assumption of normality. Consequently, we use Wilcoxon Signed-Rank tests, a nonparametric test, to determine if:

$$I_{j(i)}(\alpha_i, T_i, D_i, \theta_A) = I_{j(i)}(\alpha_i, T_i, D_i, \theta_C) \quad (17)$$

and

$$A_{j(i)}(\alpha_i, T_i, D_i, \theta_A) = A_{j(i)}(\alpha_i, T_i, D_i, \theta_C) \quad (18)$$

Where

- $i$ is a researcher
- $j$ is a project of researcher $i$
- $\alpha_i$ is the capabilities of researcher $i$
- $T_j$ is the time period when project $j$ is researched
$D_j$ is the field of research of project $j$
$\theta_A$ is an academic institution
$\theta_C$ is an commercial institution

A Wilcoxon Signed-Rank test is a nonparametric alternative to the paired $t$ test. The Wilcoxon Signed-Rank test statistic, $W$, approaches a normal distribution as the $n$ pairs of observations increase. Since $n$ is greater than 25 in our study, $W$ approximates a normal distribution. For more information on the Wilcoxon Signed-Rank test please refer to the Appendix. We test to determine if $I$ and $A$ in an academic environment are different from $I$ and $A$ in a commercial environment. We hypothesize that the importance and appropriability scores for Stanford and non-Stanford patents are equal, as presented by equations (17) and (18). In using this test, we hold $\alpha$, $T_i$, and $D_i$ fixed. $\alpha$, is held fixed by using a within-subjects design. $T_i$, is held fixed by extending the within-subjects design to included five year time periods in which Stanford and non-Stanford research is compared. Finally, $D_i$ is held fixed by performing the test separately for each research area.

For a complete list of descriptive statistics by research area please see Appendix Table A. Table 2 and Table 3 present the results of the Wilcoxon Signed-Rank test according to research area for importance, $I$, and appropriability, $A$.

<table>
<thead>
<tr>
<th>Research Area</th>
<th>W</th>
<th>E(W)</th>
<th>Standard Error of W†</th>
<th>Z Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td>7101</td>
<td>13514</td>
<td>3053.8</td>
<td>-2.1**</td>
</tr>
<tr>
<td>Computers &amp; Electronic</td>
<td>27847.0</td>
<td>58685</td>
<td>5055.4</td>
<td>-6.1**</td>
</tr>
<tr>
<td>Drugs &amp; Medical</td>
<td>8715.0</td>
<td>29670</td>
<td>4365.6</td>
<td>-4.8**</td>
</tr>
<tr>
<td>Mechanical</td>
<td>2079.0</td>
<td>4795.5</td>
<td>679.1</td>
<td>-4.0**</td>
</tr>
<tr>
<td>Other</td>
<td>632.5</td>
<td>4128</td>
<td>521.7</td>
<td>-6.7**</td>
</tr>
</tbody>
</table>

*Significant at 0.1 level.
**Significant at 0.05 level.
† Standard Errors of W are adjusted for ties.
The results in Table 2 illustrate that the researcher’s work environment does have a significant impact on the importance of a research project but not necessarily on the appropriability of a project. The direction of this significant difference was unexpected. From Appendix Table A, we see that non-Stanford importance scores are consistently higher than Stanford importance scores. Table 3 shows that the appropriability of a research project is only significantly different in the research areas of Chemical and Other. The direction of the difference for Chemical is also unexpected. In the Chemical research area, Stanford patents have a higher appropriability score than non-Stanford patents.

5.1.2 Interpretation of Results of Institution Hypothesis Testing

The importance level of Stanford research was consistently less important than non-Stanford research regardless of research area. Although significant differences exist between Stanford and non-Stanford patents, the direction of these differences is unexpected according to our behavioral model. We did not expect that Stanford research would be less important than non-Stanford research. These findings are contrary to those of Trajtenberg, Henderson, and Jaffe (1997) who found that the importance of university patents is actually higher than commercial patents in aggregate.

For an explanation of this finding we examine the assumptions used to employ the importance measure. At a conceptual level, the notion of importance as measured by patents is
similar to the notion of importance as measured by academic publications. A patent citation represents a prior disclosure that limits the scope of a claim or presents the state of current technologies (Trajtenberg, 1990). This fundamental meaning of a patent citation parallels an academic publication citation. There is an important distinction, however, between the type of research represented by a patent and an academic publication. A patent represents research that has commercial use while an academic publication represents research that is not likely to have direct commercial applications. As a result, our measure of importance is successful at gauging importance in a commercial context but not necessarily in a scholarly context. Hence, for the university researcher, an important patent—a highly cited patent—is not the same as an important publication—a highly cited publication.

These differences in research that lead to an important patent and an important academic publication encourage the division of labor between academic and industry settings, as explained by Rosenberg and Nelson (1994). A university researcher values producing important academic work and is skillful in this pursuit. The university researcher does not necessarily value producing important commercial technology and as a consequence does not have highly cited patents. An industry researcher is more skillful at and more highly values producing important commercial research. Consequently, industry researchers will likely have more highly cited patents than university researchers.

The results of Trajtenberg, Henderson, and Jaffe (1997) may be the result of aggregating patent data across all universities. At universities that have less of an academic research orientation than Stanford, more utility may be derived from an important patent. For researchers at these other universities, an important patent may be more highly valued than at Stanford. Hence, importance scores for patents at these universities would likely be higher than importance
scores for patents from a commercial setting. It seems that differences in patent valuation may explain the findings of this study in light of Trajtenberg, Henderson, and Jaffe (1997).

Although importance varies with institution for all research areas, appropriability does not. The appropriability level of research did not differ significantly with work environment. This indistinguishable response of a researcher to a change with the work environment suggests that the utility functions of a researcher in an academic and in a commercial environment behave the same way with regard to appropriability.

This finding may be explained by the OTL’s administration of the patent process causing a selection bias. The OTL does not pursue a patent on any research unless it has potential licensees already lined up. Hence, any patent that the OTL does pursue will likely have substantial market value similar to that of commercial patents. Consequently, the appropriability levels for patents within Stanford and outside of Stanford are indistinguishable due to the OTL selection bias.

The OTL screening of a patent’s appropriability may also cause a de facto screening of a patent’s importance. Important patents represent more fundamental technologies. A fundamental technology may not be as easily marketed by the OTL as more applied, narrowly focused technology. Hence, the OTL selection bias may also explain our findings with regard to importance.

In Chemical research, we find that the appropriability scores of Stanford patents are higher than those of non-Stanford patents. The unexpected direction of this significant difference can be explained by factors unique to the Chemical industry. Campbell and Landau (1998) explain, “In the chemical industry generally, excluding pharmaceuticals, patent protection has not been dominant…trade secrets are ever present” (p. 188). Since trade secrets are the
main mechanism used to protect valuable inventions rather than patents, patents in Chemical industry are likely to have low appropriability scores. In contrast, legal protection is sought on the majority of university research that has commercial value. Hence, university patents are likely to have higher appropriability scores than industry patents. Consequently, the significant different between Stanford and non-Stanford patents and its direction are due to factors specific to this research area.

5.1.3 Evaluation of Institution Hypothesis Assumptions

In testing the effect of work environment on the importance and appropriability of a project we assume that a researcher’s preferences for research given her capabilities, \( \alpha \), do not change over time. The implications of the institutional hypothesis results may change if we relax this assumption. For importance, the effect of a significant difference between Stanford and non-Stanford patents is not entirely attributable to the work environment of the researcher, \( \theta_{E(i)} \). For example, earlier in her career, an academic researcher may favor producing fame maximizing research. Later, however, if the researcher begins working in a commercial environment, her preferences may change so that she focuses on producing research that is of particular interest to her but is neither fame nor fortune maximizing. This latter preference is referred to by Graff, Heiman, and Zilberman (2002) as “love of the hunt” (p. 92). It is likely that we controlled for shifting capabilities of a researcher through \( T \), unless the preferences changed rapidly. If preferences do change rapidly, the significant differences between research environments are only partially attributable to work environment.

Pursuing a research agenda of interest rather than one that also considers fame or fortune also has implications for our findings regarding appropriability. A commercial researcher’s preferences may begin to change rapidly because he may not experience the full benefits of
creating research valued by the marketplace. In response, the commercial researcher, like the academic researcher may focus on research that has is less appropriable but has more personal interest. As in the importance analysis, it is likely such changes in preference are controlled for by using a within-subjects design and including time dummies.

5.2.1 Effect of Environment of Research Characteristics, Sponsorship Hypothesis

In investigating the sponsorship hypothesis, we examine the effect that different sources of funding have on characteristics of academic research. The character of supported research may be influenced by the differing objectives of project sponsors. To determine the effect that sponsorship has on research, we use the following statistical models:

\[ I_j = B(\theta_{S(i)} + T_j + D_j + S_i) \]  \hspace{1cm} (19)

\[ A_j = \Gamma(\theta_{S(i)} + T_j + D_j + S_i) \]  \hspace{1cm} (20)

\[ G_j = \Omega(\theta_{S(i)} + T_j + D_j + S_i) \]  \hspace{1cm} (21)

\[ L_j = \Psi(\theta_{S(i)} + T_j + D_j + S_i) \]  \hspace{1cm} (22)

Where

i is a researcher
j is a project of researcher i
\( \theta_{S(i)} \) is the sponsor of project j
Tj is the time period when project j is researched
Dj is the field of research of project j
Si is the school of researcher i
Ij is the importance of a project j
Aj is the appropriability of a project j
Gj is the generality of a project j
Lj is the licensing revenue of a project j

To test the sponsorship hypothesis, we again use patents as a proxy for research products. The sample consists only of Stanford patents. It is important to note that we cannot statistical control for a researcher’s capabilities, \( \alpha \), in this test. As described earlier, we create quantitative
characteristic scores for the level of importance, appropriability, and generality of a patent using citations measures. We then categorize patents according to the type of sponsorship the patent received. The sponsorship categories are constructed according to institution. Any research that is funded only by Stanford University is grouped under Stanford. Every category listed in Table 4 receives funding from two sources except the Stanford category. In addition to the listed sponsor, each category receives support from Stanford since the research is completed through the use of Stanford facilities. For example, all Government sponsored research is also supported by Stanford since Stanford maintains the research equipment. Table 5 presents the mean of each measure according to sponsor category.

Table 4: Stanford Patent Counts for Each Measure according to Sponsor Category

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>Patents Citation Measures (Generality, Importance, Self-Citation)</th>
<th>Licensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford</td>
<td>182</td>
<td>76</td>
</tr>
<tr>
<td>Private</td>
<td>159</td>
<td>106</td>
</tr>
<tr>
<td>Government</td>
<td>511</td>
<td>329</td>
</tr>
<tr>
<td>Non-Profit</td>
<td>28</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5: Mean for Each Measure according to Sponsor Category

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>Generality</th>
<th>Importance</th>
<th>Self-Citation</th>
<th>Licensing (millions $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford</td>
<td>.22</td>
<td>4.26</td>
<td>.09</td>
<td>0.25</td>
</tr>
<tr>
<td>Private</td>
<td>.32</td>
<td>7.73</td>
<td>.12</td>
<td>.15</td>
</tr>
<tr>
<td>Government</td>
<td>.21</td>
<td>5.02</td>
<td>.08</td>
<td>3.75</td>
</tr>
<tr>
<td>Non-Profit</td>
<td>.19</td>
<td>2.79</td>
<td>.08</td>
<td>.60</td>
</tr>
</tbody>
</table>

To determine if sponsorship does have an effect on research characteristics, we conduct a 3-way ANOVA procedure. A 3-way ANOVA procedure is employed because categorical group averages are being compared while controlling for 3 factors. The main benefit of a 3-way ANOVA is being able to hold two factors fixed while investigating the effect of the third. The three factors in our ANOVA procedure are the sponsor categories, the researcher’s school, and
the technological category. These factors are treated as random effects. The research characteristics of Generality, Importance, Self-Citations, and Licensing Revenues are the dependent variables. We test if different sources of funding affect any of these research characteristics. We hypothesize that research characteristics will change with sponsorship. For a complete results of the ANOVA testing please refer to Appendix Table B.

From Appendix Table B, we find that the only significant differences are for Generality across technological categories, $cat$, and in the interaction of researcher’s school and technological category, $school*cat$. More importantly, we find no significant differences reported across any of the sponsorship categories, $sponsor$, for the four dependent variables. Furthermore, sponsorship categories are additive in all interactions, meaning the difference between levels of $sponsor$ and the levels of other factors are not significantly different.

5.2.2 Interpretation of Results of Sponsorship Hypothesis Testing

From our ANOVA testing, we find that sponsorship does not have a significant effect on the behavior of a researcher. The major implication of the sponsorship hypothesis results is that research is not influence by different sources of funding. Furthermore, internally funded research has the same qualities of importance and appropriability as externally funded research. This suggests that having a relationship with an industry sponsor does not change the incentives structure faced by an academic researcher enough to alter the character of his work.

Comparing the significance test results for importance and appropriability in the sponsorship hypothesis to the results in the institution hypothesis illustrates the strength of the response of researchers to differences in environmental incentive structures. For importance, the change in environment from commercial to academic had a significant impact on the research’s level of importance. In contrast, within the academic environment, working with
different sponsoring organizations does not have an impact on the level of importance. The differing response levels of importance suggest that the incentives at an institutional level are a great deal more influential than the incentives at a interaction level between academic labs and the objectives of outside sponsors. For appropriability, as in the institution hypothesis, a researcher’s behavior did not change in response to changes in environmental incentives. This consistent behavior suggests that a researcher’s preference for developing approbicable work is carried with the researcher; it is embedded in $\alpha$, the capabilities of the researcher.

5.2.3 Evaluation of Sponsorship Hypothesis Assumptions

To investigate the sponsorship hypothesis, we assume that the capabilities of a researcher were independent of the sponsorship he received. If the result of the hypothesis testing was significant relaxing this assumption would have made it difficult to determine the causation from tastes, $\alpha$, and sponsorship, $\theta_{s(i)}$. Because the results of the testing are not significant, relaxing this assumption would likely not have a noteworthy effect on the behavioral interpretation of the testing results. The result is still that the characteristics of a research project are not influenced by the type of sponsorship it receives.

Although sponsorship is determined not to have a significant influence over the research characteristics of Stanford professors this result cannot be necessarily scaled to all academic researchers. One of the primary reasons that Stanford researcher are not affected by different sources of funding is due to their unique position. Most Stanford professors have reached a high degree of prestige in their profession. This fame has given the professors substantially more bargaining power than the majority of academic researchers. A Stanford professor is in the position to choose and negotiate with the sponsors of their research. Consequently, sponsorship
may have an impact in other universities where the researchers are not so uniquely positioned and empowered.

5.3.1 Career Anticipatory Hypothesis

The final aspect of this study evaluates the career anticipatory behavior of a researcher. The career anticipatory hypothesis tests if a researcher engages in tool building activity. Building from our behavioral model, the appropriability level for research of the entrepreneurial professor and non-entrepreneurial professor can be described by the following statistical model:

\[ A_{j(E)} = (\alpha_E + D_j + \theta_E) \]  
\[ A_{j(NE)} = (\alpha_{NE} + D_j + \theta_{NE}) \]

Where

\( j \) is a project of a researcher  
\( A_{j(E)} \) is the appropriability of a project \( j \) for the entrepreneurial researcher  
\( A_{j(NE)} \) is the appropriability of a project \( j \) for the non-entrepreneurial researcher  
\( \alpha_E \) is the capabilities for the entrepreneurial researcher  
\( \alpha_{NE} \) is the capabilities for the non-entrepreneurial researcher  
\( D_j \) is the field of research of project \( j \)  
\( \theta_E \) is the anticipated work environment of the entrepreneurial researcher  
\( \theta_{NE} \) is anticipated work environment of the non-entrepreneurial researcher

From this model, we determine if \( A \) for a project varies depending on if a professor is an entrepreneur or not. A professor can be an entrepreneur due to either his preference for a given set of capabilities, \( \alpha \), or his anticipated work environment, \( \theta \).

It is important to note that for a researcher to participate in tool building activities, his work outside of Stanford must be in the same research area as his Stanford projects. This category is defined by variable \( \text{cat} \), which is a grouping of the narrower USPTO patent classification scheme. If tool building occurs, the appropriability levels for Stanford projects of non-entrepreneurs should be less than or those of entrepreneurs:
\[ A_{j(E)}(\alpha_E, D_j, \theta_E) > A_{j(NE)}(\alpha_{NE}, D_j, \theta_{NE}) \]  

(25)

Where

- \( j \) is a project of a researcher
- \( A_{j(E)} \) is the appropriability of a project \( j \) for the entrepreneurial researcher
- \( A_{j(NE)} \) is the appropriability of a project \( j \) for the non-entrepreneurial researcher
- \( \alpha_E \) is the capabilities for the entrepreneurial researcher
- \( \alpha_{NE} \) is the capabilities for the non-entrepreneurial researcher
- \( D_j \) is the field of research of project \( j \)
- \( \theta_E \) is the anticipated work environment of the entrepreneurial researcher
- \( \theta_{NE} \) is anticipated work environment of the non-entrepreneurial researcher

To investigate the career anticipation hypothesis, we use the patent portfolio of Stanford researchers. Those researchers that have assigned patents outside of Stanford are entrepreneurial researchers. Researchers who only have patents assigned to Stanford are non-entrepreneurial researchers. A third group of researchers, those that have never assigned patents to Stanford but have assigned patents outside of Stanford, is not included in this testing. It is not possible to observe if tool building takes place for this third group. Hence, our sample consists only of the first two groups: researchers who have Stanford and non-Stanford patents and researchers who have only Stanford patents. Excluding the third group of researchers, biases our results against tool building behavior.

To test the career anticipation hypothesis we use a nonparametric test, the Mann-Whitney U test, because we again cannot make the assumption of normality. The Mann-Whitney U test is an alternative to the two-sample \( t \) test and is only slightly weaker. When both sample populations have greater than 10 cases, the Mann-Whitney U statistic approaches a normal distribution. For more information on Mann-Whitney U statistic please refer to the Appendix.

We use the test to determine if \( A \) for an entrepreneurial researcher’s Stanford projects is significantly different from \( A \) for a non-entrepreneurial researcher’s Stanford projects. In employing this test we hold the research area, \( D_j \), of the researcher fixed by separately
performing a test for each research area. We hypothesize that $A$ for an entrepreneurial researcher’s Stanford projects is greater than $A$ for a non-entrepreneurial researcher’s Stanford projects as described in equation (25).

For a complete list of descriptive statistics for appropriability scores of entrepreneurs and non-entrepreneurs please see Appendix Table C. Table 6 presents the results of the Mann-Whitney U tests on appropriability according to research area.

| Table 6: Results of Mann-Whitney U Test by Research Area |
|-----------------|-----------------|-----------------|-----------------|
|                 | U               | E(U)            | Standard Error  | Z-statistic     |
| Chemical        | 1841.5          | 1636.4          | 227.9           | -.9             |
| Computers & Electronic | 6414.0      | 5294.2          | 622.1           | -1.8*           |
| Drugs & Medical | 3161.5          | 2325.0          | 363.7           | -2.3**          |
| Mechanical      | 345.0           | 288.7           | 70.4            | -.8             |
| Other           | 52.0            | 24.6            | 22.8            | -1.2            |

*Significant at 0.1 level.
**Significant at 0.05 level.

The results in Table 6 illustrate that there is a significant different in appropriability levels for projects in the areas of Computers & Electronics and Drugs & Medical. While the research areas of Chemical and Mechanical do not have significant differences between appropriability scores for Stanford and non-Stanford research.

5.3.2 Interpretation of Results for Commercial Awareness Hypothesis Testing

The significant differences in the Computers & Electronic and Drugs & Medical categories indicate that the appropriability scores for research produced at Stanford by entrepreneurial researchers are higher than for research produced by non-entrepreneurial researchers. From our behavioral model, this implies that tool building does occur and that the entrepreneurial researcher anticipates the future market value of her field of research. Her anticipatory behavior permits her to utilize these tools developed while an academic to receive pecuniary returns outside of academia.
Comparing the career anticipatory hypothesis to the sponsorship hypothesis, gives us a unique conceptual framework to understand tool building. For the entrepreneur, future commercial returns and support for research is analogous to a source of sponsorship for research. The entrepreneur is responding to a future source of funding by participating in tool building. A possible reason an entrepreneur responds to future funding rather than current sponsored funding is that future funding provides a substantially larger incentive. There are two likely explanations why this incentive is significantly larger. First, unlike in the sponsorship hypothesis, the entrepreneur does not necessarily have any bargaining power to influence their future commercial sponsor. Consequently, she must produce research valued by the future sponsor. Second, the entrepreneurial researcher may overestimate the future pecuniary return to her research. This over estimation of future rewards may also cause a researcher to significantly change her behavior and thereby characteristics of her research.

In addition to comparing these results to the sponsorship hypothesis, we considered them in light of the institution hypothesis results. Comparing the results of these two hypotheses illuminates a behavioral phenomenon among Stanford researchers. In the case of Drugs & Medical and Computers & Electronic, we found that institution did not have an effect on the appropriability level of a research. This lack of a significant difference between the academic and commercial environments may be due to tool building. If researchers in these categories were already producing appropriable research as academics, it is unlikely there is a dramatic difference between their academic research and their commercial research in terms of appropriability. A second explanation for this result is that the individual academic researcher values appropriable research more than the researcher at a corporation. The individual academic researcher needs to produce appropriable research to become a successful entrepreneur. An
industry researcher does not necessarily have the equal pressure to produce appropriable research unless they also plan to become entrepreneurs. The industry researcher, however, does need to produce research that is appropriable. Hence, the industry researcher produces research that has low appropriability relative to industry and the entrepreneurial academic researcher produces research that has high appropriability relative to academia. When compared to each other, however, we do not observe a significant difference. Both of these explanations suggest that we did not find differences in appropriability across institution because of entrepreneurial behavior.

In the area of Chemical research, we found that institution did have an effect on the appropriability score of research and also that researchers do not participate in tool building. From the nature of the Chemical industry as explained by Campbell and Landau (1998), a patent is probably not an accurate representation of a Chemical researcher’s entrepreneurial behavior. The lack of patenting in the Chemical industry is supported by findings from Hall, Jaffe, and Trajtenberg (2001). The authors find a steady decline in Chemical patents over the last 3 decades. Since patenting is declining and trade secrets are an important aspect of this industry as identified by Campbell and Landau (1998), it is likely that other mechanisms such as consulting are more widely used by entrepreneurial researchers. Consequently, tool building is likely occurring but we do not observe it.

In the area of Mechanical research, we also find that researchers do not participate in tool building as measured by patents. Hall, Jaffe, and Trajtenberg (2001) find a decline in Mechanical patents that is similar to the decline Chemical patents. It seems patents also fail to capture the entrepreneurial behavior of Mechanical researchers. As in the Chemical research area, another mechanism such as consulting is likely to support the entrepreneurial behavior of researchers in this field.
The results of the career anticipatory hypothesis should also be evaluated considering the ambient environment around Stanford University. Within the last three decades Silicon Valley has been extremely supportive of entrepreneurship and particularly encouraging in the Computers & Electronic and Drugs & Medical research areas. This unique environment could also contribute to our finding of tool building occurring in the Computers & Electronic and Drugs & Medical but not in the Chemical and Mechanical research areas.

5.3.3 Evaluation of Career Anticipatory Assumptions

When investigating the career anticipation hypothesis, we extended the methodology of Hall, Jaffe, and Trajtenberg (2001) to group the research areas of Stanford projects into five sub-fields: Chemical, Computers & Electronic, Drugs & Medical, Mechanical and Others. In grouping projects into these larger sub-fields, we assume the research in these areas is similar in nature. If we relax this assumption, the results of the career anticipatory hypothesis may change. If we group research projects according to their narrow 3-digit patent classification scheme, we would likely find tool building to occur only in sectors of the Drugs & Medical and Computers & Electronic research areas. We would also find that other sectors of these research areas have no observable tool building behavior. This implies certain sectors of these industries are more prone to encourage tool building than others. Entrepreneurs would gravitate to these tool building sectors opposed to the non-tool building sectors where the barriers to entrepreneurship are greater. It seems unlikely that many entrepreneurs emerge from the non-tool building sectors. The entrepreneur is presumably either drawn to tool building sectors or becomes an entrepreneur by working in these sectors. Consequently, sectors of these larger research areas facilitate entrepreneurial behavior because of the factors inherent to the research.
A second aspect of the career anticipatory hypothesis that is important to note is a limitation of the data. Within the data, the entrepreneurial researcher who is a “true entrepreneur” is not observed. The data used in this study tracks the patent portfolios for all researcher who have ever disclosed to the OTL. The “true entrepreneurial researcher” would never disclose to the OTL. Consequently, the behavior of these researchers is not observed. This bias, however, makes the findings of the commercial awareness hypothesis even more revealing. The commercial awareness hypothesis testing results are currently biased towards zero. If we were able to observe the behavior of the “true entrepreneurial researcher”, our current findings would be the same but even more statistically significant.

6 CONCLUSION

The purpose of this study is to develop a richer understanding of the behavior of university professors. In particular, we set out to answer two questions: first, what is the effect of different institutional incentive structures on a researcher’s behavior and second, does an researcher exhibit career anticipatory behavior due to future incentives? To answer these questions, we tested three hypotheses which investigated institutional effects on research characteristics, sponsorship effects on research characteristics, and the behavior of entrepreneurial professors. In addition to the substantive results of our testing, our study has produced useful methodology insights.

Our unique methodological procedures are possible due to a merged dataset which combined the NBER patent data file constructed by Hall, Jaffe, and Trajtenberg (2001) with proprietary information obtained from Stanford’s OTL. By merging these two datasets, we created individual patent portfolios for Stanford University researchers. These patent portfolios allowed us to gain insights into the research behavior of academic researchers over their careers.
In addition, the nature of our individualized researcher data also enabled us to use a within-subjects experimental design in portions of our empirical analysis. A within-subjects experimental design helps alleviate the common problem of unobserved heterogeneity across subjects by holding factors unique to an individual researcher fixed. Both of these benefits were obtained without the considerable time and resource burden typically associated with behavioral studies that rely on survey data. Consequently, the methodology developed in this study should easily apply to similar studies of other universities.

In addition to the methodology insights of our study, substantive results were found that contribute to our understanding of a university researcher’s behavior in the context of university-industry relationships. From the three hypotheses tested, two factors seem to influence the behavior of a researcher: the researcher’s capabilities, $\alpha$, and the researcher’s work environment, $\theta$. In evaluating the findings of this study, we should again note the type of research we observe in a patent. A patent represents research that is typically more commercially oriented than most academic research. As discussed earlier, a patent may not be an accurate representation of the importance of research in a scholarly sense but does provide an excellent measure of the appropriability of research.

From our empirical analysis, we learned that the appropriability of a research project is not entirely a function of work environment. For the university researcher, the appropriability of his research is more likely determined by the researcher being an entrepreneur. Although different factors encourage a researcher to become an entrepreneur, once an entrepreneur, the appropriability of a researcher’s work is affected. In particular, we find that a researcher participates in tool building—producing commercial oriented research—if she is an entrepreneur.
The appropriability results for each of our hypotheses illuminates an important aspect of technical advance within the context of the university-industry relationship. Many industry-university relationship studies have neglected to consider the behavior of the university researcher. The findings of this study suggest that an important aspect of this relationship is related to the incentives faced by the academic. The university researcher has a distinct set of factors he is maximizing—fame, fortune, and freedom. These factors influence how he reacts to incentives in his work environment and to his own capabilities.

Considering these factors within the context of the sponsorship debate, we find that funding source does not have a significant influence over a researcher. The main fears of university officials are that university-industry collaborations compromise the “primary educational and research missions [of universities]” as well as “the university as a credible and impartial resource” (Bowie, 1994, p. 161). This concern that a researcher’s values and scholarship are corrupted by corporate money is not supported by our findings. Corporate support and relationships do not directly encourage a researcher to alter his academic values or change the nature of his work. A researcher’s work changes to become more commercial only with the more substantial influence of anticipating entry into industry.

The anticipatory behavior of a researcher to enter into the private sector is a commonly overlooked mechanism for technology transfer from universities. The most frequently relied on form of technology transfer is the licensing of university developed technologies to industry; a system primarily initiated and encouraged by the Bayh-Dole Act. Mowery and colleagues (2001) explain that the Bayh-Dole Act of 1980 encouraged the “negotiation of exclusive licenses between universities and industrial firms for the results of federally funded research” (p. 102). In addition to these imposed channels of knowledge transfer, another possible channel for transfer
exists in the university researcher himself. From the career anticipatory hypothesis, we observe
that tool building occurs among entrepreneurial researchers. Eventually researchers that
participate in tool building become heavily involved with industry. The migration of researchers
from university to industry is another way to provide knowledge transfer from universities. The
foremost benefit of such knowledge transfers is that a much more substantial set of knowledge
moves with the person as compared to the knowledge that moves with more traditional
mechanism like patents or publications. A problem with this knowledge transfer mechanism,
however, is it could lead to understaffing at universities.

A number of future investigations could be conducted in this growing area of economics
in view of this study’s results. Foremost of these studies is a verification of our behavioral
model which would substitute our patent data with more traditional income and academic
publication data. Building on this study’s findings regarding tool building behavior, future
research should explore more precisely the events that lead to tool building and the barriers faced
by entrepreneurial professors who attempt to commercialize their research. A third area of
research should also be pursued to examine the differences between knowledge transfer via a
migration mechanism versus traditional publication mechanisms.

The findings of this study provide a perspective on the behavior of university researchers
within the context of the university-industry relationship. Our results indicate that the university
professor is not as easily corrupted as some may fear. Additionally, our research suggests the
migration of researchers from university to industry is an important mechanism for knowledge
transfer. Consequently, future examinations of American universities’ interaction with industry
should also consider the behavior of university professors.
APPENDIX

Statistical Testing

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a nonparametric alternative to the paired-observations $t$ test. Unlike the Mann-Whitney test, the Wilcoxon Signed-Rank test accounts for the magnitude differences between paired values not just their signs by considering ranks of differences. We assume that the distribution of the differences of the two populations is symmetric, mutually independent, and the measurement scale is at least an interval (Aczel, 1999).

The Wilcoxon $W$ statistic is defined as the smaller of two sums of ranks:

$$ W = \min \left( \sum (+), \sum (-) \right) $$

Where

$\sum (+)$ is the sum of the positive ranks

$\sum (-)$ is the sum of the negative ranks

The population mean of $W$ is:

$$ E(W) = \frac{n(n+1)}{4} $$

Where

$n$ is the sample size of the paired observations

The standard deviation of $W$ is:

$$ \sigma_w = \sqrt{\frac{n(n+1)(2n+1)}{24}} $$

Where

$n$ is the sample size of the paired observations
The standardize $z$ statistic is:

$$z = \frac{W - E(W)}{\sigma_W}$$

**Mann-Whitney U Test**

The Mann-Whitney U Test, also known as the Wilcoxon rank sum test, is a test of equality of two populations means. It is a nonparametric alternative to the two-sample $t$ test. In fact, the Mann-Whitney U test is only slightly weaker than the $t$ test. To use the test, we must make two assumptions. First, we assume that the samples are random. Second, we assume that the samples are drawn independent of each other (Aczel, 1999).

The Mann-Whitney U statistic is:

$$U = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1$$

Where

- $n_1$ is the sample size from population 1
- $n_2$ is the sample size from population 2
- $R_1$ is the minimum sum of the ranks of the sample populations

The population mean of the distribution of $U$ is:

$$E(U) = \frac{n_1 n_2}{2}$$

Where

- $n_1$ is the sample size from population 1
- $n_2$ is the sample size from population 2

The standard deviation of $U$ is:

$$\sigma_u = \sqrt{n_1 n_2 (n_1 + n_2 + 1) / 12}$$
Where

\( n_1 \) is the sample size from population 1
\( n_2 \) is the sample size from population 2

The standardized \( z \) test statistic is:

\[
\begin{align*}
    z &= \frac{U - E(U)}{\sigma_U} \\
\end{align*}
\]
### Additional Tables

**Table A: Descriptive Statistics for Stanford and Non-Stanford Patents according to Research Area for Institution Hypothesis**

<table>
<thead>
<tr>
<th>Research Area</th>
<th>Importance Score</th>
<th></th>
<th>Appropriability Score</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stanford</td>
<td>Non-Stanford</td>
<td>Stanford</td>
<td>Non-Stanford</td>
</tr>
<tr>
<td>Chemical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.1</td>
<td>3.5</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>Median</td>
<td>.5</td>
<td>1.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.2</td>
<td>5.6</td>
<td>.20</td>
<td>.09</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>33</td>
<td>35</td>
<td>1</td>
<td>.73</td>
</tr>
<tr>
<td>Computers &amp; Electronic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.5</td>
<td>5.3</td>
<td>.05</td>
<td>.04</td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>7.2</td>
<td>7.9</td>
<td>.15</td>
<td>.09</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>75</td>
<td>85</td>
<td>.75</td>
<td>1</td>
</tr>
<tr>
<td>Drugs &amp; Medical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.9</td>
<td>3.2</td>
<td>.06</td>
<td>.04</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>5.3</td>
<td>8.6</td>
<td>.18</td>
<td>.11</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>32</td>
<td>129</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mechanical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.5</td>
<td>4.3</td>
<td>.065</td>
<td>.055</td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>3.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.2</td>
<td>5.3</td>
<td>.17</td>
<td>.11</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>22</td>
<td>26</td>
<td>1</td>
<td>0.67</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.1</td>
<td>4.0</td>
<td>.015</td>
<td>.041</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.3</td>
<td>4.5</td>
<td>.09</td>
<td>.08</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>22</td>
<td>31</td>
<td>1</td>
<td>.42</td>
</tr>
</tbody>
</table>
Table B: Results of 3-way ANOVA for Sponsorship Hypothesis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>F-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School</strong></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>.2</td>
</tr>
<tr>
<td>Importance</td>
<td>.8</td>
</tr>
<tr>
<td>Appropriability</td>
<td>.4</td>
</tr>
<tr>
<td>Licensing Revenues</td>
<td>0</td>
</tr>
<tr>
<td><strong>Cat</strong></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>2.6**</td>
</tr>
<tr>
<td>Importance</td>
<td>1.3</td>
</tr>
<tr>
<td>Appropriability</td>
<td>1.1</td>
</tr>
<tr>
<td>Licensing Revenues</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sponsor</strong></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>1.4</td>
</tr>
<tr>
<td>Importance</td>
<td>.9</td>
</tr>
<tr>
<td>Appropriability</td>
<td>.4</td>
</tr>
<tr>
<td>Licensing Revenues</td>
<td>0</td>
</tr>
<tr>
<td><strong>School*Cat</strong></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>1.7**</td>
</tr>
<tr>
<td>Importance</td>
<td>.7</td>
</tr>
<tr>
<td>Appropriability</td>
<td>.8</td>
</tr>
<tr>
<td>Licensing Revenues</td>
<td>0</td>
</tr>
<tr>
<td><strong>Cat*Sponsor</strong></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>1.0</td>
</tr>
<tr>
<td>Importance</td>
<td>1.0</td>
</tr>
<tr>
<td>Appropriability</td>
<td>1.1</td>
</tr>
<tr>
<td>Licensing Revenues</td>
<td>0</td>
</tr>
<tr>
<td><strong>School<em>Cat</em>Sponsor</strong></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>.7</td>
</tr>
<tr>
<td>Importance</td>
<td>.9</td>
</tr>
<tr>
<td>Appropriability</td>
<td>.3</td>
</tr>
<tr>
<td>Licensing Revenues</td>
<td>0</td>
</tr>
</tbody>
</table>

*Significant at 0.1 level.
**Significant at 0.05 level.
### Table C: Appropriability Descriptive Statistics for Entrepreneur and Non-Entrepreneur according to Research Area for Career Anticipatory Hypothesis

<table>
<thead>
<tr>
<th>Research Area</th>
<th>Entrepreneur</th>
<th>Non-Entrepreneur</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.25</td>
<td>.20</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Computers &amp; Electronic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.0873</td>
<td>.0525</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.18</td>
<td>.11</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>.5</td>
</tr>
<tr>
<td><strong>Drugs &amp; Medical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.1111</td>
<td>.0323</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.25</td>
<td>.09</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>.4</td>
</tr>
<tr>
<td><strong>Mechanical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.1316</td>
<td>.0683</td>
</tr>
<tr>
<td>Median</td>
<td>.03</td>
<td>0</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.24</td>
<td>.1</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>.25</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.0450</td>
<td>.0714</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>.07</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.17</td>
<td>.08</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
<td>.14</td>
</tr>
</tbody>
</table>
References


