Undergraduate Economic Review

Volume 4 | Issue 1 Article 11

2008

Analysis of the Relationship Between Research Grants and Medical Patents: Are the Number of Medical Patents Dependent on NIH and NSF Funding

Jay H. Shah
Union College

Analysis of the Relationship Between Research Grants and Medical Patents: Are the Number of Medical Patents Dependent on NIH and NSF Funding

Abstract

The United States invests more into basic medical research than any other country. This research, mostly done at medical schools, leads to medical technology patented by private firms through a phenomenon known as knowledge spillover. In light of this relationship, this thesis will investigate the relationship between the public investments in the basic medical research sector and the resulting effect on medical patents, contingent on geographic localization with state-level data. It is hypothesized that the larger the investments awarded to basic medical research organizations, the greater the number of medical patents within the state. Quantifying this relationship has important implications for modeling change, economic growth and science policies.

The proposed model relates the number of medical patents per year in a given state, the dependent variable, as a function of the following independent variables: National Institutes of Health (NIH) funding, National Science Foundation (NSF) funding, the number of civilian scientists and engineers within the state, and a state-specific research productivity parameter. The model used is similar to Jaffe (1989) but differs as this relationship focuses on different independent variables. The data come from a variety of sources including Jaffe and Trajtenberg (2002), the NIH, and the NSF. These data points will be used to run different regression analyses to test the relationship in question.

With the regression analysis and time lag adjustments, it is anticipated that the number of patents per year in a given state will be positively correlated with the number of scientist and engineers and the amount of medical research funding allotted to each state. The performed regressions support this hypothesis and resulted in statistically significant coefficients. This relationship will also evaluate the economic aspect of medical research and determine the effect of knowledge spillover from medical institutions to nearby private inventors. Ultimately, this may allow for more precise institutional and

geographical allocation of research investments for the purpose of achieving more medical innovations, thus advancing the field of medicine as a whole.

ANALYSIS OF THE RELATIONSHIP BETWEEN RESEARCH GRANTS AND MEDICAL PATENTS: ARE THE NUMBER OF MEDICAL PATENTS DEPENDENT ON NIH AND NSF FUNDING?

BY:

JAY H. SHAH

* * * * * * * * *

Submitted in partial fulfillment
of the requirements for
Honors in the Department of Economics

UNION COLLEGE
MARCH, 2008

ABSTRACT

SHAH, JAY H. Analysis of the Relationship Between Research Grants and Medical Patents: Are the Number of Medical Patents Dependent on NIH and NSF Funding? Department of Economics, March 2008.

The United States invests more into basic medical research than any other country. This research, mostly done at medical schools, leads to medical technology patented by private firms through a phenomenon known as knowledge spillover. In light of this relationship, this thesis will investigate the relationship between the public investments in the basic medical research sector and the resulting effect on medical patents, contingent on geographic localization with state-level data. It is hypothesized that the larger the investments awarded to basic medical research organizations, the greater the number of medical patents within the state. Quantifying this relationship has important implications for modeling change, economic growth and science policies.

The proposed model relates the number of medical patents per year in a given state, the dependent variable, as a function of the following independent variables: National Institutes of Health (NIH) funding, National Science Foundation (NSF) funding, the number of civilian scientists and engineers within the state, and a state-specific research productivity parameter. The model used is similar to Jaffe (1989) but differs as this relationship focuses on different independent variables. The data come from a variety of sources including Jaffe and Trajtenberg (2002), the NIH, and the NSF. These data points will be used to run different regression analyses to test the relationship in question.

With the regression analysis and time lag adjustments, it is anticipated that the number of patents per year in a given state will be positively correlated with the number of scientist and engineers and the amount of medical research funding allotted to each state. The performed regressions support this hypothesis and resulted in statistically significant coefficients. This relationship will also evaluate the economic aspect of medical research and determine the effect of knowledge spillover from medical institutions to nearby private inventors. Ultimately, this may allow for more precise institutional and geographical allocation of research investments for the purpose of achieving more medical innovations, thus advancing the field of medicine as a whole.

TABLE OF CONTENTS

CHAPTER 1:

INTRODUCTION

A. Basic research vs. Applied research

Over the course of the past decade, there has been an astonishing increase in the amount of capital allocated to basic medical research organizations. The primary objective of basic research is to gain more comprehensive knowledge and understanding of general variables, not to create or invent something. Fundamental research, as it is also known, is considered exploratory and mostly driven by the researcher's or researcher's organizations' interests and intuition (Moses et al., 2005).

Basic research focuses on scientific principles and discoveries and is considered to be a precursor to developments in applied research. Applied research utilizes the discoveries of basic research to commercialize applications through patenting. The research is designed to solve practical problems of the modern world. For example, basic research allowed for the discovery of the relationship between electricity and magnetism while applied research created the invention of the integrated circuit, the key component in microprocessors. Basic research is typically funded through charitable associations and governmental agencies and performed at public institutions. In contrast, applied research is primarily funded through private organizations and conducted at private firms.

B. Knowledge spillover

In reality, however, there is a strong relationship between basic and applied research organizations in the form of knowledge spillover. Knowledge spillover is defined as the

transmission of knowledge to others beyond the intended boundaries (Fallah and Ibrahim, 2004). Within the past two decades, considerable interest in quantifying the spillover effects between the public and private research sectors has attracted more researchers. This measurement is important as the amount of knowledge spillover has an immediate impact on the firm's innovation efforts. In addition, firms may decide to free ride on other firm's research efforts if large spillovers are present. When an interviewer stated that Siemens, Europe's fifth largest enterprise in terms of employees, conducts little research and development, the CEO replied by stating, "We have installed listening posts around." This free riders approach allows firms to benefit from research conducted outside their firm. From an individual firm's perspective, spillovers may create a disincentive effect for own research efforts, however large spillovers can increase the appropriation possibilities of other firm's knowledge. Because of these possibilities, researchers have focused on measuring the transmitted knowledge between firms (Kaiser, 2002).

Knowledge spillovers which can be transformed to explicitly stated information such as a patent is called 'codified' knowledge (Kaiser, 2002). Therefore, for codified knowledge measurements, most researchers rely heavily on patent and patent citation data to observe the effect of geographic localization of knowledge spillover. Most researchers strongly suggest that knowledge spillovers are geographically localized and statistically significant.

The knowledge spillover effect of medical research has received little attention. The research, whether basic or applied, is conducted to add to the overall knowledge in the field of medicine. Advancements in the medical field are responsible for significant benefits, such as vaccines and treatments, which result in increased longevity. While immense amounts of capital are invested into basic medical research, the size of the knowledge spillover to the private sector is still unknown. Estimating this relationship will allow for quantifying and further understanding the gains from publicly funded medical research.

C. Contribution and hypothesis of this study

In light of the medical research knowledge spillover effect, this thesis will measure the relationship between the investments made into the basic medical research sector and the resulting effect on medical-based patents, contingent on geographic localization within states. The idea behind this relationship is that public and private medical schools receive funding from governmental agencies, such as the National Institutes of Health (NIH), and utilize the funding for basic medical research. Due to knowledge spillovers, private firms and inventors use this information to conduct applied research measurable by the granting of medical patents. Observing the link between basic medical research and medical patenting allows for measuring the return to public funds spent on medical research. It is hypothesized that the larger the investments awarded to the medical sector, the greater the number of medical patents on a state-wide basis. Observing this relationship has important implications for modeling change, economic growth and science policies. It will also allow for more precise institutional and geographical

allocation of research grants for the purpose of achieving more medical innovations, thus advancing the field of medicine as a whole.

CHAPTER 2:

LITERATURE REVIEW

A. Recent trends in medical research

According to Moses et al. (2005), annual U.S. spending on medical research has doubled in the past decade from $37.1 billion in 1994 to more than $94 billion in 2003, an increase of 102%. Dr. Moses and colleagues determined this value by including the major sponsors of medical research: federal government, state and local governments, private not-for-profit entities including foundations, and the industry itself. Their research found that of the $94 million spent in 2003, 40% came from the public sector while 60% was funded privately. Relative proportions from all public and private sources have remained constant. This parallel growth of public and private spending suggests a strong interrelationship between the sources of funds and their use. The public sector consists of federal, state and local funding while the private sector includes private not-for-profit companies and industry support (pharmaceutical, biotechnology, and medical device firms). Moses et al. (2005) also report that the total medical research expenditures at universities and colleges were $19.6 billion for 2002, up from $10.7 billion in 1995. The Federal sector accounted for 64% of expenditures while institutional funds were responsible for the next largest share at 17%. Institutional funds include subsidy from physician practice income, endowments, and hospitals' support of research.

Moses et al. (2005) reports that of the $94 billion spending for medical research, the National Institutes of Health (NIH) and pharmaceutical companies contribute $26.4 billion, 28% of the total, and $27.0 billion, 29% of the total, respectively. Biotechnology

companies contribute $17.9 billion, about 19% of the total spending while medical device companies account for $9.2 billion, 10% of the total. The NIH is the largest federal funder of medical research. In 2005 alone, the NIH spent over $11.5 billion of their $28 billion budget in NIH awards to medical schools (Cech, 2005).

The medical research industry is divided into three sectors: drug, biotechnology and medical device research. Of these, medical device researchers saw funds rise 264% in the past decade. In contrast, support for pharmaceutical innovations and biotechnology firms grew 89% and 98% respectively. A study done by Reuters (2005) reports that the additional dollars spent on medical research has yielded disappointing results with regards to pharmaceutical companies. There has been a significant lag in the development of new, useful drugs compared to research results produced by medical device and biotechnology industries.

In response to the article by Moses et al. (2005), Woolley and Propst (2005) conducted a study and reported on the public mood to such an increase in research spending. A compilation of 70 state surveys and 18 national surveys found that more than half of Americans think that the U.S. should spend more on research while two-thirds of all Americans are willing to increase taxes to achieve this goal. A very high proportion of Americans (94%) say that medical and health research is important to the US economy, and 79% agree that basic science research should be supported by the federal government, "even if it brings no immediate benefits."

Based on the vast amount of investments reported, the analysis of the relationship between grants for basic medical research and the rate of applied research patents is necessary to understand the effects of the allocated capital for research purposes. This topic must be further analyzed to gain better understanding of the social rate of return for grants given to the basic medical research sector.

B. Flow of university knowledge

Many scholars warn that quantifying the actual flow of knowledge is impossible because the flow is invisible and leaves no paper trail that can be measured or tracked (Krugman, 1991). A thorough study by Griliches (1990) established quite a strong relationship between research and development and the number of patents received at both the cross-sectional level and the time-series dimension. Griliches (1990) states that a change in a firm's research and development expenditures are paralleled by changes in patent numbers. As a result of this study, most economists studying this topic utilize patents and patent citations to measure innovation.

Beginning with one of the most well-known economists in this area, Adam Jaffe (1993) determined that although knowledge spillovers may be invisible, the geographically localized effect is quite large and significant. Jaffe (1989) studied the existence of geographically mediated spillovers from university research to commercial innovation, an example of movement from basic to applied research. He used state-level time-series data on corporate patents, corporate research and development (R&D), and university

research. Based on his findings, university research directly affected local innovation by inducing industrial R&D spending.

Jaffe (1989) stated that there is even more reason, in comparison to other basic research organizations, to believe that spillovers exist from universities to firms, since universities have less incentive to try to keep research secrets. He examined the patents granted to private companies by state over time and regressed these on industry R&D and university research. He found that the analysis of state-level corporate patent activity provides some evidence of the importance of geographically mediated commercial spillovers from university research (Jaffe, 1989).

The indirect effect of university research is also important; there appears to be an association between industry R&D and university research. The causality is that university research results in an increase in industry R&D and thus, a state that improves its university research system will increase local innovation by both attracting industrial R&D as well as increasing its productivity (Jaffe, 1989).

C. Clusters and geographic localization

Similar to Jaffe's (1989) approach on industry research and development, Koo (2005) observed geographical patterns of patents in knowledge-based industry clusters. These clusters consist of interconnected businesses within the same geographic areas that bind together for the purpose of increased productivity. The data provide evidence that as businesses are closer together, there are increased knowledge spillovers and subsequently, increased patenting. Knowledge-based industry clusters can stimulate

inter-firm and inter-industry knowledge spillovers thereby making firms more innovative. Examples of such geographic clusters are Silicon Valley in California and the Research Triangle in North Carolina. These areas consist of interrelated businesses. Silicon Valley continues to be the leading high-tech hub because of its large number of engineers and venture capitalists, while the Research Triangle Park is home to numerous high-tech companies and enterprises with one of the highest concentrations of Ph.Ds per capita (U.S. Census, 2000).

Economists have been long arguing that clustering contributes to economic growth. Even before Koo (2005), Grossman and Helpman (1994) centered the theory of endogenous economic growth on the idea that accumulation of basic knowledge will eventually find a way to productive applications and therefore lead to economic growth. This perspective hints to the premise that location plays a role in knowledge spillovers. Grossman and Helpman (1992) explain that the rate of economic growth could vary across regions due to spillovers causing an increase in local knowledge creation as affected by geography. Basically, certain regions have higher rates of growth if the clustering is closer together.

Similar to the idea of proximity in clusters by Grossman and Helpman (1992), Keller (2002) observes the benefits from spillovers diminish with distance. Using global data that encompasses most of the innovative activity throughout the world, he created a model to estimate that technology is to a substantial degree local opposed to global. He estimated that the strength of spillovers is halved at about 1,200 kilometers. Although Keller performed the study internationally, a similar principle applies within just the U.S.

For some varying distance, dependent on the technology, the benefits of the spillover effect are halved. As the firms get closer within a cluster, more knowledge spillover can be expected. Relating this topic to the paper, the smaller the distance between the medical institution and the private institution, the higher the likelihood university spillover in the form of patenting will occur.

D. Time and quality effects

A few studies have focused on explicitly introducing time to measure the rate of diffusion on the effects of university patenting. Jaffe and Trajtenberg (1996) developed a model to evaluate subsequent patent citations in light of viewing knowledge diffusion. They examined the citation probability ratio by technological field and found the drugs and medical field to have the highest lag, on average six years, for predicted citation frequency. Citation probability ratio is the probability that a patent will cite a past patent to properly provide reference. This was greater than all other fields, including chemicals, electronics, and mechanical, which were all around four years. The conjecture to higher lag times may be attributed to the long lead times associated with approval from the Federal Drug Administration (FDA). This finding is consistent with most of the research done by medical institutions. Once a patent is granted, for any use of the patent, the drug or medical device must go through FDA approval, a process that contains many steps and can be very lengthy. Because of this, the lag in years is greatest for these types of patents.

The university patenting rate has exploded over the past forty years. This phenomenon is unusual since universities are dedicated to the widespread dissemination of their research rather than self-patenting. Henderson et al. (1998) report that university patents are both more important and more general than the average patent although this trend is declining. Henderson et al. (1998) show that these two traits of university patents have fallen at the same time as the sheer number of university patents have increased. The decrease in importance and generality is largely a result of a very rapid increase in the number of low-quality patents. This increase in the number of patents is attributed to an increased rate of technology transfer to the private sector.

While the private sector is filing higher-quality patents, universities are patenting basic, low-quality patents for medical usages that have probably increased the social rate of return to university research. According to Henderson et al. (1998), the majority of the economic benefits of university research come from inventions in the private sector. Regardless, as a result of the increased number of university patents there may be an increase in the propensity to patent. The rate of knowledge transfer from university to private sector could also increase (Henderson et al., 1998).

These articles support and provide a solid encompassing background in the relationship of the knowledge spillover effect. While some present a vague relationship between universities and patents, none analyze the specific relationship between grants into the basic medical research sector and the resulting medical patents from the private sector.

CHAPTER 3:

THEORETICAL MODEL

A. Rival vs. Excludable goods

The goal of this thesis is to explore the grants made into basic medical research and the subsequent spillover into the private sector as an output of medically-related patents. The observation was limited to a state-wide analysis. Basic scientific research is an example of a public good that is both non-rival and non-excludable. Non-rival goods allow many people to consume the same good at the same time without reducing the consumption of the others while non-excludable goods are goods that are impossible to prevent people who have not paid for it from enjoying its benefits. Basic research cannot be privately provided or traded in competitive markets. All researchers can take advantage of the research at the same time. In contrast, applied research is excludable and non-rivalrous (Romer, 1990). This difference is important in understanding knowledge spillover. It is known that basic research is conducted at universities and medical institutions which spillover into applied research at private firms and ultimately leads to patenting (Jaffe (1989), Henderson et al. (1994)). In this knowledge spillover relationship between medical patents and funding, it is hypothesized that the amount of money allocated to the medical research sector through grants has a direct and positive correlation with the number of assigned medical patents, on a state-level analysis.

B. Included variables

In lieu of observing the funding into the medical sector, two of the largest sources of federal funding were included: National Institutes of Health (NIH) and National Science Foundation (NSF) (Moses et al., 2005). The idea is that the increase in funding from

these sources allows for an increase in the medical research done at the college/university receiving the funding. As more research is performed, there will be more knowledge spillover into the private sector that ultimately leads to an increase in medical patents for that geographic location. At the NIH and NSF, funding is allocated to specific organizations (colleges, businesses, fellowships) based on submitted proposals. The funding is not general but rather specific. Therefore, once the funding is received, the organization begins working on their proposed plan.

Submitted proposals to the NIH and NSF are typically initiated by qualified scientists and engineers and officially submitted by their employing organization. Including the names of the scientists and engineers on the proposal provides it with credibility. Because of this relationship and influence, it is believed that the number of scientists and engineers within a certain area have a direct relationship with the rate of medical patenting for that area. Research-based companies are fueled by a large and high-caliber workforce of scientists and engineers and are considered to be the key engines of growth (The Progressive Policy Institute, 2002).

In an effort to control the state's previous economic growth and success prior to 1990, a variable was included to support the dependent variable of medical patents. The variable consists of all the recorded utility patents from the initial recording year of 1963 through 1990, for each state. Using this theoretical background and explained inputs, the following data sets were used in estimating the empirical model.

CHAPTER 4:

THE DATA

To analyze the relationship between medical patents and investments made to the medical research sector, data were compiled from a few different sources. The data for the dependent variable, medical patents, were obtained from a book by Jaffe and Trajtenberg (2002). The data included as a compact disc along with the book were originally taken from the United States Patent and Trademark Office (USPTO) with certain computations added by Jaffe and Trajtenberg. The data file includes all utility patents from the USPTO's TAF computerized database, granted during the period 1963 to December 1999. Each observation contains classification information reflective of the U.S. Patent Classification System as of December 31, 1999. The database contains approximately 3 million utility patent observations, each with detailed information on the innovation itself, the technological area to which it belongs, the inventors and the organization to which the inventors assigned the patent property right. Of these observations, over 70,500 patents were identified in the Drugs and Medical category and Surgery & Medical Instruments subcategory. For these observations, the inventor's application state was also included. The data were compiled by application year to obtain the number of medical patents per state by year. In addition to the medical patent observations, a time invariant variable represented a base of all patents, by state, from the years 1963-1990. This state-level variable, *Pat1990*, was included as a regressor to proxy for differences in local knowledge base and innovativeness across states.

For the 123 medical institutions nationwide, the National Institutes of Health (NIH) Awards to Medical Schools based on rank have been obtained from 1990 to 2005 from

the NIH. The data were given in both Total Awards (TA) as well as Research Grants (RG) for each medical school however the correlation between them was high. The Total Awards value was used in the regressions. The amounts were then aggregated based on state to conform to the rest of the data.

Following the NIH grants, the National Science Foundation (NSF) provided the value of awards for each state from 1990 to 2005. The NSF funds research and education in science and engineering through grants, contracts, and cooperative agreements. The Foundation accounts for about 20 percent of federal support to academic institutions for basic research.

The number of scientists and engineers by state and year were obtained through the Division of Science Resources Statistics (SRS) of the National Science Foundation (NSF). This division publishes the Science and Engineering State Profiles. Although these data are compiled biennially, values for the unreported years were estimated as an average of the year before and after.

Due to the limiting data set of both the NIH and NSF, the earliest year for observations is 1990. The patent data, however, begins in 1963 and ends in 1999. These limiting data sets only allow for complete observations from 1990-1998 for a total of 450 observations organized by state and year. Descriptive statistics for each variable are shown in Table 1 (appendix).

CHAPTER 5:

EMPIRICAL MODEL

A. OLS regression

The OLS model used to estimate the knowledge spillover from investments into the medical sector to private patenting was based on the framework initially articulated by Zvi Griliches (1979) and implemented by Ariel Pakes and Griliches (1984), Jaffe (1986), Jaffe (1989), and others. It represents the framework of a knowledge production function as a modified Cobb-Douglas model with four inputs:

$$\textbf{(1)} \quad \log(\boldsymbol{MedPatent})_{it} = \beta_0 + \beta_1 \log(\boldsymbol{SE})_{i(t-l)} + \beta_2 \log(\boldsymbol{NIH})_{i(t-l)} + \beta_3 \log(\boldsymbol{NSF})_{i(t-l)} + \beta_4 \log(\boldsymbol{Pat1990})_i + \varepsilon_{it}$$

where i indexes the unit of observation (states), t is the year year, l is the lag structure, *MedPatent* is the number of utility patents under the Drugs and Medical category and Surgery & Medical Instruments subcategory, *SE* is the number of scientists and engineers, *NIH* and *NSF* is the amount of grants awarded for each organization respectively, and *Pat1990* is the base number of patents achieved by state from 1963-1990. The ε represents the stochastic error term.

Equation 1 models the governing relationship between medical patents per state and a function of scientists and engineers, investments from the NIH and NSF, and a proxy for differences in the local knowledge base and innovativeness across states. The *SE* variable serves as a measure for private resources. Both the *NIH* and *NSF* variables measure the total funding, in thousands of dollars, per state. The inclusion of NSF funding also helps to capture the effect of inter/intra-industry spillover. Lastly, the *Pat1990* variable serves as a state-specific research productivity parameter. The natural

log of each variable in the equation has been taken which not only accounts for diminishing returns but also turns the model into a modified Cobb-Douglas production function. Therefore, each variable's coefficient is the output elasticity which measures the responsiveness of the medical patents to changes in levels of inputs used in the production. The Cobb-Douglas production function also allows for determining constant returns to scale. If all coefficients are summed and they are equal to one, the production function has constant returns to scale. A Wald Test was completed to test constant returns to scale.

Once a medical school/university receives funding for a specific project, the research is furthered prior to the spillover of knowledge. The time period prior to spillover can vary from a year to three or four, depending on the technology and stage of research. In addition, the nature of patent applications often requires time and effort to establish the claims associated with the patent application. These factors draw attention to the need of a proper lag structure. It is somewhat unreasonable to believe that funding given to a medical school will allow for furthering of research to the point that the resulting knowledge spillover assists applied research and then translates into a patent application, all within one year. Therefore, Equation 1 contains the variable l to adjust for different lag structures.

Jaffe (1989), however, uses a contemporaneous regression, meaning a lag structure was not implemented in his equation. While Jaffe is a well-known and respected researcher in this field, his lack of lag within his model is incoherent. To believe that research in universities both spillover and translate into a patent application within the same year

may not be unreasonable. From personal experiences with knowledge spillover, it required about a little less than two years to progress from receiving funds for college research to produce and submit the application for a medical patent. Using Equation 1, various lag structures may be applied to allow different observations of the knowledge spillover effect.

B. Fixed effects specification regression and inclusion of state-specific research productivity parameter

As with any economic relationship, there are many independent inputs that affect the dependent variable. Within this relationship possible variables unique to each state include but are not limited to: the industry mix (Malerba and Montobbio, 2003), local demand, cultural factures and level of urbanization (Audretsch and Lehmann, 2005). Because the proposed model cannot include all the possible variables, two approaches can be taken. The first, state fixed-effects model, can be broken down into cross-section and period specifications. With a state fixed-effects model as shown in Equation 2, each state has its own intercept. In practice this means that the coefficients are estimated using information about how each variable varies over time.

$$\textbf{(2)} \quad \log(\textit{MedPatent})_{it} = \beta_{0i} + \beta_1 \log(\textit{SE})_{i(t-l)} + \beta_2 \log(\textit{NIH})_{i(t-l)} + \beta_3 \log(\textit{NSF})_{i(t-l)} + \beta_4 \log(\textit{Pat1990})_i + \varepsilon_{it}$$

Therefore, no information about the cross-state variation in the variables can be used to estimate the coefficient. The period fixed-effects estimation imposes time independent effects for each state and attempts to control for unobserved heterogeneity. Both cross-section and period fixed-effects are examined through regressions.

The other approach to adjust for unaccounted variables is to include a regressor that is a proxy for differences in local knowledge base and innovativeness across states. *Pat1990*, the sum of all patents from 1963-1990 for each state, is representative of a state-specific research productivity parameter.

C. Two-Stage Least Squares (2SLS) regression

A possible problem with the proposed model is endogeneity, which is when the value of one independent variable is dependent on the value of other predictor variables. Endogeneity may result in significant correlation between the unobserved variables contributing to both the independent and dependent variables. This can also result in biased estimators, incorrect regression coefficients, and correlation with the error term (Dowd and Town, 2002). Additionally, the correlation between the dependent variables can create significant multicollinearity violating the assumptions of a standard regression model. To correct for this problem, an instrumental variable regression within a two-stage least squares (2SLS) can replace the standard OLS regression. The equation for this approach is represented in Equation 3.

$$\textbf{(3)} \quad \log(\textit{MedPatent})_{it} = \beta_0 + \beta_1 \log(\textit{SE})_{i(t-1)} + \beta_2 \log(\textit{NIH})_{i(t-1)} + \beta_3 \log(\textit{NSF})_{i(t-1)} + \beta_4 \log(\textit{Pat1990})_i + \varepsilon_{it}$$

Above each of the possible endogenous variables, which are all of them, bars (or hats) above the variables would represent estimated values instead of actual values. Because all the inputs within the equation are possible endogenous variables, an instrument panel with all the variables, including different lags, was included.

CHAPTER 6:

ESTIMATION RESULTS

A. OLS regressions

As stated above, the regressions were performed with data of 9 years, 1990-1998, for 50-states. The panel least squares regressions in Table 2 display the coefficients of the independent variables on the dependent variable, *logMedPat*. The values in parentheses are t-statistics. As this is a Cobb-Douglas production function, each variable's coefficient is the output elasticity of that input which measures the responsiveness of the dependent medical patents as a change in levels of inputs used in the production, ceteris paribus. From regression 1 in Table 2, it can be inferred that if the number of scientists and engineers doubles within a state, the number of medical patents for that state will increase by 76% while doubling NSF and NIH grants causes a 12% and 25% increase respectively. The coefficient for *logNSF* is significant at the 5% level with *logSE* and *logNIH* significant at 1%. The R-squared is 0.774 with a Durbin-Watson statistic of 0.622.

As discussed above, money invested into the medical research sector in one year may not directly result in a patent application in that year. A lag of one year of all independent variables is used in regression 2 of Table 2. All coefficients are statistically significant at the 5% level. This lag of one year means that the number of scientists and engineers and NIH and NSF investments in 1992 are related to the number of medical patent applications in 1993. Two, three and four year lagged regressions with the similar coefficients are displayed in Table 2 as regression 3, 4 and 5 respectively. At the three and four year lagged time frames, the coefficient of the NSF variable is not significant

past the 10% level. As the number of lagged years increase, the coefficients fluctuate minimally however the R-squared value continually decreases. Also, each additional lag reduces the number of observations by 50.

To check for returns to scale in these regressions, the coefficients of each variable were summed. The total came to 1.2, which represents increasing returns to scale as it is greater than one. To verify, a Wald Coefficient Restrictions Test was performed to compute a test statistic based on the unrestricted regression. The Wald statistic measures how close the unrestricted estimates come to satisfying the restrictions under the null hypothesis. If the restrictions are in fact true, then the unrestricted estimates should come close to satisfying the restrictions. With a coefficient restriction of $c(1) + c(2) + c(3) = 1$, the probability of the f-statistic (0.0005) and chi-squared (0.0004) values are both significant at the 1% level. These results mean that there is increasing returns to scales and they are significant. For all other regressions, regression 1 will serve as a reference point.

B. Fixed effects specification regressions

Regression 6 in Table 3 estimates a fixed effects model. Fixed effects specifications, either period-fixed or cross-section fixed, is a familiar approach of removing specific means from the dependent variable as well as exogenous regressors. After they are removed, the specified regressions are performed on the demeaned data. Regression 6 shows that with a period and cross-section fixed effect specification all of the coefficients lose significance, along with very different coefficients. The high R-squared value, 0.961 can be attributed to the fact that each state received its own intercept while no corss-state

variation exists. To compare one state to another, regressions 7-12 only have fixed period effects. Regression 7 shows 1% significance for the scientists and engineers' variable, the NIH and the NSF. The coefficients are also similar to those found in regression 1. Regressions 8-12 demonstrate consecutive increases in lag starting with a one year lag in regression 8. As the regressions are lagged, the effect of *SE* is reduced 10% by regression 12 (5 year lag) relative to a contemporaneous function. The coefficients for the NSF and NIH variables increase as the years lagged increase. All three variables remain significant at the 1% level throughout the five year lag. The R-squared remains around 0.839 throughout these regressions.

C. Research productivity parameter inclusion regressions

Table 4, presents the results from regressions 13-18, all contain various lags in an OLS regression model while including the specific research productivity parameter, *logPat1990*. Regression 13 shows no lag. The addition of the proxy variable caused the coefficient of scientists and engineers to decrease from 0.76 (from regression 1) to 0.33 however all variable coefficients remain statically significant at the 1% level. This decrease represents a possible correlation between these variables. The presence of scientists and engineers may be representing a previously innovative state. Scientists and engineers may move to a state with a higher level of innovation. The addition of the *logPat1990* also caused this regression to represent constant returns to scale.

Regressions 14, 15, 16, and 17 display the results from a 1, 2, 3 and 4 year lag respectively. The one year lag shows a 50% decrease in the coefficient of *logPat1990*. The following year, the value recovers back to 0.321. The three year lag, regression 16,

shows all coefficients to be statistically significant at the 5% level with similar values to regression 13. Regression 17, a lag of 4 years, produces similar coefficients however loses the significance of NSF funding at the 10% level. Regression 18 demonstrates a lag of four years with a fixed-period effect specification. This causes all variables to be statistically significant at the 5% level and produces an R-squared of 0.848. The coefficient for NSF grants also exceeds the coefficient for NIH funding.

D. Two-Stage Least Squares regressions

As mentioned before, these regressions all contain possible endogenous regressors which can cause bias OLS coefficients. For example, the addition of the base *logPat1990* variable caused the coefficient of the scientists and engineers to decrease substantially. This could mean that instead of an increase in the number of scientists and engineers leading to an increase in medical patents, scientists and engineers could be moving to more innovative states because of past success or increased NSF grants. To adjust for these endogenous regressors, a two-stage least squares (2SLS) analysis was performed. To correlate endogenous variables, the 2SLS contains an instrument list. Within the regressions, the instrument list consisted of each of the variables included in the model with corresponding lags: *logSE*, *logNSF, logNIH, logPat1990*. The two stages refer to (1) a stage in which new dependent variables are created to substitute for the original ones, and (2) a stage in which the regression is computed in OLS fashion, but using the newly estimated variables.

Table 5, regressions 19-24, display the results from two-stage least squares regressions with different lagged instrument panels. Regression 19 of Table 5 shows a regression

with a contemporaneous lag including the state-level variable *logPat1990* as part of the instrument list and as a regressor. All coefficients remained statistically significant at the 1 % level with an overall R-squared value of 0.791. All coefficient values except *SE* are almost identical to those of regression 1. A significant drop from 0.760 to 0.334 was noticed for the *SE* variable. A one year lag shown in regression 20 of Table 5 causes the scientists and engineering variable to lose overall significance as the coefficient also decreases. This is noticed throughout the two, three and four year lags. Regressions 21, 22, and 23 display two, three and four year lags respectively. The coefficient for NSF loses significance at the four year level while the NIH variable and *logPat1990* remained statistically significant at the 1% level. The final regression, regression 24, applied a fixed period effect to the four year lagged two stage least squares regression. All three variables, except *SE* are statically significant at the 5% level however the coefficient for NSF grants exceeds that of the NIH.

CHAPTER 7:

CONCLUSIONS

A. Summary of findings

The analysis of state-level medical patent activity as a function of inputs from the NIH, NSF, and the number of scientists and engineers provides evidence of the importance of knowledge spillover from the basic to applied sectors of medical research. Most of the regressions demonstrated statistically significant coefficients for the independent variables, even after accounting for various lags. Controlling for the state-specific research productivity parameter through 2SLS resulted in a strong correlation between the number of medical patents per state and the amount of NIH funding allotted as well as the base research productivity parameter. Controlling for period fixed-effects, 5% significance was observed for all independent variables except the scientists and engineers and also caused the coefficient for NSF to exceed that of the NIH. The different purpose of NSF and NIH grants initially rejects this finding as the NIH awards are specifically for medical school research which leads to more patented medical devices while the NSF is in the general field of science. However, it is possible that awards from the NSF require more years to translate into research that can spillover into the applied research sector and ultimately lead to medical patents.

In most of the regressions, a low Durbin-Watson (between 0.42 and 0.70) statistic was present. The Durbin-Watson statistic test is used to detect the presence of autocorrelation in the residuals from the regression analysis. A value around two indicates that there appears to be no autocorrelation. If the Durbin-Watson statistic is substantially less than

2, there is evidence of positive serial correlation within the error term which may cause biased coefficients. A future study should address the possible serial correlation.

These regression analyses approached from OLS, fixed-effects, and two-stage least squares methods, mostly concluded consistent and statistically significant results that yielded similar coefficients for the NIH and NSF variables. On average, if the amount of NIH and NSF funding were to double, the number of medical patents per state would increase by around 13% and 25% respectively. On average, before *logPat1990* was included, doubling the number of scientists and engineers per state would cause roughly a 70% increase in medical patents. Once the specific research productivity parameter *logPat1990* was included, the *SE* term dropped to around 30% indicating possible correlation. This evidence supports the proposed hypothesis.

B. Policy Implications

Understanding this relationship is important for future policy implications commonly drawn from new growth theory. It states that as a result of convexities in knowledge production and the resultant increasing returns, knowledge resources, such as research and development should be publicly supported. These spillovers may also serve as a focus for enhancing public policy to continue economic growth and development (Audretsch and Feldman, 2003).

Loscalzo (2006) states that we have recently entered another period of stagnant funding for the NIH. Having doubled between 1998 and 2003, the NIH budget is expected to be $28.6 billion for fiscal year 2007, a 3.8 percent decrease which is the first budgeted reduction in NIH support since 1970. This downturn is more severe than any previously

faced and threatens to erode the benefits of previous investments. It takes many years for institutions to develop scientists and engineers skilled in modern research techniques and to build the costly, complicated infrastructure necessary for medical research. Loscalzo (2006) goes on to state that the situation is unlikely to improve anytime soon as the resources required by the war and erosion of the tax base by the current administration's fiscal policies are expected to have long-term, far-reaching effects.

Studies observing the benefits of NIH funding, such as this thesis, can show the positive effects of NIH funding in not only basic research but also show that the research actually translates into significant medical patents within the applied research sector. If enough studies are produced with significant results, the NIH could provide substantial evidence to the government to hopefully increase funding to continue medical research.

C. Suggestions for Future Research

Based on the low Durbin-Watson statistic, future research should focus on increasing this value by controlling for serial correlation. To adjust for serial correlation, more inputs could be included in the regressions to account for possible variables unique to each state. In addition, a larger data set including more recent data could provide more information about the relationship between medical patents and public funding over time.

REFERENCES:

"AAMC: Medical Schools." 2007. Association of American Medical Colleges. <http://www.aamc.org/medicalschools.htm>.

Audretsch, D B., and E E. Lehmann. Does the Knowledge Spillover Theory of Entrepreneurship Hold for Regions? Elsevier, 2005.

Audretsch, David, and Maryann P. Feldman. "Knowledge Spillovers and the Geography of Innovation." Handbook of Urban and Regional Economics 4 (2003).

Cech, Thomas R. "Fostering Innovation and Discovery in Biomedical Research." Journal of the American Medical Association 294 (2005): 1390-1393.

"Census 2000 Gateway." U.S. Census Bureau. U.S. Government. 2 Mar. 2008 <http://www.census.gov/main/www/cen2000.html>.

Dowd, Bryan, and Robert Town. Does X Really Cause Y? Washington: AcademyHealth, Changes in Health Care Financing and Organization, 2002.

Fallah, M.H. & Ibrahim, S. 2004. Knowledge Spillover and Innovation in Technological Clusters. Proceedings, IAMOT 2004 Conference, Washington, D.C., April 2004.

Griliches, Zvi. "Issues in Assessing the Contribution of Research and Development to Productivity." The Bell Journal of Economics 10 (1979): 92-116.

Griliches, Zvi. "Patent Statistices as Economic Indicators: a Survey." Journal of Ecnomic Literature 28 (1990): 1661-1707.

Grossman, Gene M., and Elhanan Helpman. "Endogenous Innovation in the Theory of Growth." The Journal of Economic Perspectives 8.1 (1994): 23-44.

Grossman, Gene M., and Elhanan Helpman. Innovation and Growth in the Global Economy. Cambridge: MIT P, 1991.

Henderson, Rebecca, and Iain Cockburn. "Measuring Competence? Exploring Firm Effects in Pharmaceutical Research." Strategic Management Journal 15 (1994): 63-84.

Jaffe, Adam B., "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents Profits and Market Value," American Economic Review, December 1986, 76, 984-1001.

Jaffe, Adam B. "Real Effects of Academic Research." The American Economic Review 79 (1989): 957-970.

Jaffe, Adam B., Trajtenberg, Manuel; and Henderson, Rebecca. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." Quarterly Journal of Economics 108 (August 1993): 577–98.

Jaffe, Adam B., and Manuel Trajtenberg. "Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations over Time and Across Institutional and Geographic Boundaries." Proceedings of the National Academy of Sciences 93 (1996): 12671-12677.

Jaffe, Adam B., and Manuel Trajtenberg. Patents, Citations, and Innovations: a Window on the Knowledge Economy. Cambridge: The MIT P, 2002.

Kaiser, Ulrich. "Measuring Knowledge Spillovers in Manufacturing and Services: an Empirical Assessment of Alternative Approaches." Research Policy 31 (2002): 125-144.

Keller, Wolfgang. "Geographic Localization of International Technology Diffusion." The American Economic Review 92 (2002): 120.

Koo, Jun. "Knowledge-Based Industry Clusters: Evidenced by Geographical Patterns of Patents in Manufacturing." Urban Studies 42 (2005): 1487-1505.

Krugman, P., (1991), *Geography and Trade*, (MIT Press: Cambridge).

Loscalzo, Joseph. "The NIH Budget and the Future of Biomedical Research." The New England Journal of Medicine 354 (2006): 1665-1668.

Malerba, Franco, and Fabio Montobbio. "Exploring Factors Affecting International Technological Specialization: the Role of Knowledge Flows and the Structure of Innovative Activity." Journal of Evolutionary Economics 13 (2003): 411-414.

Moses, Hamilton III; E. Ray Dorsey; David H. M. Matheson; Samuel O. Thier Financial Anatomy of Biomedical Research Journal of the American Medical Association 2005 294: 1333-1342

Romer, Paul M. "Endogenous Technological Change." The Journal of Political Economy, 98 (1990): s71-s102.

The Progressive Policy Institute (PPI). Technology, Innovation, and New Economy Project. The 2002 State New Economy Index. June 2002. <http://www.neweconomyindex.org/states/2002/index.html>.

Toole, Andrew A., and Dirk Czarnitzki. "Biomedical Academic Entrepreneurship Through the SBIR Program." Journal of Economic Behavior & Organization 63 (2006): 716-738.

Wallsten, Scott J. "The Effects of Government-Industry R&D Programs on Private R&D: the Case of the Small Business Innovation Research Program." The RAND Journal of Economics 31 (2000): 82-100.

Woolley, Mary, and Stacie M. Propst. "Public Attitudes and Perceptions About Health-Related Research." Journal of the American Medical Association 294 (2005): 1380-1384.

U.S. Department of Health & Human Services. National Institutes of Health. Office of Extramural Research. NIH Extramural Awards Ranking Tables Medical Schools Current Data. 2 Nov. 2006. <http://grants1.nih.gov/grants/award/>.

"US Medical Research Spending Rises, Results Lag." Reuters Sept. 2005. <http://www.redorbit.com/news/health/246290/us_medical_research_spending_rises_results_lag/ >.

APPENDIX:

Table 1: *Descriptive statistics for all variables included in regressions*				
	SE	**NIH** (millions)	**NSF** (millions)	**PAT1990**
Mean	10372.85	81.74	52.43	25123.50
Median	6197.50	34.18	20.57	10281.00
Maximum	83042.00	655.58	565.20	162969.00
Minimum	801.00	0.00	0.01	549.00
Std. Dev.	12417.12	115.76	81.30	34738.82
Observations	450	450	450	450

Table 2: *This table demonstrates five regressions performed with various lags in an OLS regression model*					
	Regression 1	**Regression 2 (-1)**	**Regression 3 (-2)**	**Regression 4 (-3)**	**Regression 5 (-4)**
logSE	0.760*** (8.539)	0.751*** (7.815)	0.726*** (6.985)	0.745*** (6.557)	0.759*** (5.601)
logNSF	0.124** (2.158)	0.129** (2.096)	0.149** (2.267)	0.111 (1.567)	0.091 (0.967)
logNIH	0.249*** (6.491)	0.252*** (6.042)	0.261*** (5.736)	0.277*** (5.446)	0.091*** (4.809)
N	450	400	350	300	250
R-squared	0.774	0.765	0.761	0.749	0.735
Durbin-Watson	0.622	0.624	0.624	0.651	0.718

Note: The values in the table represent the coefficients for each independent variable. The respective t-statistic is presented in parentheses. Years lagged are in parentheses after the regression number.
*** $p<0.01$; ** $p<0.05$; * $p<0.1$;

Table 3: *This table demonstrates seven regressions performed with various lags using different fixed effects models*							
	Regression 6	**Regression 7**	**Regression 8 (-1)**	**Regression 9 (-2)**	**Regression 10 (-3)**	**Regression 11 (-4)**	**Regression 12 (-5)**
logSE	0.261 (0.862)	0.719*** (9.391)	0.715*** (8.779)	0.687*** (7.958)	0.691*** (7.443)	0.611*** (5.651)	0.616*** (5.140)
logNSF	0.003 (0.065)	0.180*** (3.577)	0.182*** (3.438)	0.208*** (3.770)	0.184*** (3.150)	0.247*** (3.274)	0.235*** (2.917)
logNIH	0.028 (0.177)	0.251*** (7.691)	0.253*** (7.212)	0.260*** (6.942)	0.276*** (6.720)	0.280*** (6.191)	0.279*** (5.484)
Fixed Cross-Section	Cross-Section	-	-	-	-	-	-
Fixed Period	Period	Period	Period	Period	Period	Period	Period
N	450	450	400	350	300	250	200
R-squared	0.961	0.841	0.838	0.841	0.838	0.839	0.839
Durbin-Watson	1.821	0.464	0.445	0.403	0.406	0.391	0.436

Note: The values in the table represent the coefficients for each independent variable. The respective t-statistic is presented in parentheses. Years lagged are in parentheses after the regression number.

*** $p<0.01$; ** $p<0.05$; * $p<0.1$;

Table 4: *This table demonstrates six regressions performed with various lags in an OLS regression model with the specific research productivity parameter, logPat1990*

	Regression 13	**Regression 14 (-1)**	**Regression 15 (-2)**	**Regression 16 (-3)**	**Regression 17 (-4)**	**Regression 18 (-4)**
logSE	0.334*** (3.006)	0.329** (2.735)	0.316** (2.416)	0.352** (2.430)	0.390** (2.285)	0.301** (2.243)
logNSF	0.170*** (3.037)	0.173** (2.881)	0.188*** (2.934)	0.146** (2.108)	0.125 (1.357)	0.273*** (3.696)
logNIH	0.227*** (6.128)	0.229*** (5.663)	0.238*** (5.366)	0.254*** (5.101)	0.253*** (4.462)	0.260*** (5.855)
logPat1990	0.330*** (6.0539)	0.327*** (5.499)	0.321*** (4.908)	0.311*** (4.219)	0.292*** (3.456)	0.248*** (3.742)
Fixed Cross-Section	-	-	-	-	-	-
Fixed Period	-	-	-	-	-	Period
N	450	400	350	300	250	250
R-squared	0.791	0.782	0.776	0.764	0.748	0.848
Durbin-Watson	0.672	0.673	0.662	0.683	0.743	0.415

Note: The values in the table represent the coefficients for each independent variable. The respective t-statistic is presented in parentheses. Years lagged are in parentheses after the regression number.

*** $p<0.01$; ** $p<0.05$; * $p<0.1$;

Table 5: *This table demonstrates six regressions preformed with various lags and instrument panels in a two-stage least squares*

	Regression 19	**Regression 20 (-1)**	**Regression 21 (-2)**	**Regression 22 (-3)**	**Regression 23 (-4)**	**Regression 24 (-4)**
logSE	0.334*** (3.006)	0.209 (1.379)	0.150 (0.822)	0.190 (0.850)	0.358 (1.584)	0.192 (1.064)
logNSF	0.169 *** (3.037)	0.282*** (2.860)	0.346*** (2.808)	0.304* (1.938)	0.168 (1.117)	0.379*** (3.313)
logNIH	0.227*** (6.128)	0.226*** (5.545)	0.232*** (5.082)	0.244*** (4.736)	0.239*** (4.073)	0.255*** (5.376)
logPat1990	0.329 *** (6.054)	0.341*** (5.654)	0.340*** (5.0441)	0.330*** (4.335)	0.304*** (3.545)	0.272*** (3.910)
Instrument Panel	logSE logNSF logNIH logPAT1990 C	logSE(-1) logNSF(-1) logNIH(-1) logPAT1990 C	logSE(-2) logNSF(-2) logNIH(-2) logPAT1990 C	logSE(-3) logNSF(-3) logNIH(-3) logPAT1990 C	logSE(-4) logNSF(-4) logNIH(-4) logPAT1990 C	logSE(-4) logNSF(-4) logNIH(-4) logPAT1990 C
Fixed Period	-	-	-	-	-	Period
N	450	400	350	300	250	250
R-squared	0.791	0.764	0.767	0.757	0.747	0.836
Durbin-Watson	0.672	0.619	0.705	0.730	0.767	0.558

Note: The values in the table represent the coefficients for each independent variable. The respective t-statistic is presented in parentheses. Years lagged are in parentheses after the regression number.

*** $p<0.01$; ** $p<0.05$; * $p<0.1$;