AN ABSTRACT OF THE DISSERTATION OF

Yin Xia for the degree of Doctor of Philosophy in Agricultural and Resource Economics, presented on September 20, 2002.

Title: Science and Technology in Cutting-Edge Agricultural Biotechnology Research

Signature redacted for privacy.

Abstract approved: ______________________________

Steven T. Buccola

The genetic engineering made possible by the discovery of recombinant DNA has played an increasingly important role in agricultural research. The present study employs a knowledge production model to assess the efficacy of, and relationship between, basic and applied research in agricultural biotechnology, allowing for both complementarity and substitutability between these two research endeavors. Practical measures of basic and applied research outputs, and a paper trail characterizing the information flows between them, are constructed using a unique database on agricultural biotechnology patents and patent-cited scientific publications.

Results suggest university bioscience research and graduate education are mostly complements, or in some cases slight substitutes, for one another. Highly ranked universities are less efficient than are their lower-ranked counterparts in producing not only graduate students but the bioscience that is cited in agricultural biotechnology patents. University R&D expenditures have been inoptimally allocated

between post-doctoral fellows and non-post-doctoral inputs. Higher returns to R&D funding would be achieved by diverting some such funds away from non-post-doctoral inputs and toward post-doctoral fellows, and away from biology programs and toward agricultural programs.

Commercial firms' agricultural and non-agricultural (primarily pharmaceutical) research are complements to one another. Firms' propensity to patent agricultural biotechnology inventions, rather than hold them as trade secrets, has increased significantly, while their propensity to patent in non-agricultural or non-biotechnology fields has fallen. Biotechnology firms have devoted too little of their R&D expenditures to scientists and engineers and too much to non-salary inputs. Boosting biotech firms' R&D expenditures would bring only a small change in their agricultural biotechnology output but a large increase in their non-agbiotech output.

In the production of agricultural biotechnology innovations alone, basic bioscience and applied biotechnology appear always to be complementary with one another. But in the production of non-agricultural innovations, bioscience and applied technology are either complements or substitutes, depending upon the manner in which R&D expenditures are allocated. In general, choices among alternative R&D inputs greatly influence the effectiveness of R&D investments in agricultural biotechnology. Complementarity between science and technology in agriculture suggests boosting communication between basic and applied research would bring high social dividends.

Science and Technology in Cutting-Edge Agricultural Biotechnology Research

by

Yin Xia

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements of the
degree of

Doctor of Philosophy

Presented September 20, 2002
Commencement June 2003

Doctor of Philosophy dissertation of Yin Xia presented on September 20, 2002

APPROVED:

Signature redacted for privacy.

Major Professor, representing Agricultural and Resource Economics

Signature redacted for privacy.

Head of the Department of Agricultural and Resource Economics

Signature redacted for privacy.

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Signature redacted for privacy.

Yin Xia, Author

ACKNOWLEDGEMENTS

It is a great pleasure to acknowledge several individuals without whom the completion of this dissertation would not have been possible.

Foremost, I would like to express my deepest appreciation to Professor Steven Buccola for his constant guidance, encouragement, and support throughout my graduate work. His extensive knowledge and critical thinking have been my most valuable resource and have led me into the door of agricultural economics. His striving for excellence, search for the truth, and care for his students will always be an example in my future career.

I deeply thank Professor Rolf Fare for his valuable help and support during the preparation of this dissertation. Meetings and discussions with him are always productive and enjoyable. I would also like to thank Professor Munisamy Gopinath for his challenging questions and helpful suggestions on empirical estimation. Special thanks also go to Professor David Birkes for serving on my committee and providing helpful comments.

Particular gratitude is dedicated to my parents and husband, for their everlasting love, understanding, and support.

Support for this research has been provided by the National Research Initiative Competitive Grants Program, CSREES, U.S. Department of Agriculture, No. 00-35400-9142.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

Page

TABLE OF CONTENTS (CONTINUED)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

DEDICATION

To my parents Qiguang Xia and Wanzhu Chen,

and

my husband Xudong Fan,

for their constant love, understanding, and support.

Science and Technology in Cutting-Edge Agricultural Biotechnology Research

Chapter 1: Introduction

Two decades ago, it was generally thought that genetically modified agricultural products were in the distant future. Since U.S. Department of Agriculture (USDA) first permitted the Flavor Savor tomato, a new GM plant variety, to be offered directly to consumers in 1992, transgenic crops have gained increasing market share. In 2001 about 26% of the corn and 68% of the soybeans and cotton grown in the United States were transgenic varieties (Crop Production Report 2001, USDA National Agricultural Statistical Service). The new biotechnology has helped increase farm-level productivity by increasing yields and reducing input requirements, and has helped enhance food quality such as flavor, texture, shelf-life, and nutritional content. But this is only the beginning. New generations of transgenic crops and animals will embody novel product characteristics such as oil, vitamin, starch, carbohydrate, and protein content tailored for specific uses, enabling specific demand profiles to be catered to in ways unthinkable in the past. For example, a recent development in agbiotech is the design of GM agricultural products intended for industrial manufacturing, such as a new corn variety that may be used to produce plastics or inks. Despite recent consumer safety worries and environmental concerns, agricultural biotechnology is reshaping agriculture as profoundly as mechanical, biological, and chemical innovation paradigms did during the past 150 years (Zilberman, Yarkin, and Heiman, 1999).

The revolution was first brought about by significant progress in basic biological science. Discovery of the double-helix structure of DNA, and the subsequent development of recombinant DNA technology in the 1950s and 1960s, led to practical protocols for transferring potentially useful genes from one organism to another. Exploitation of those protocols soon led to patentable products, which finally became commercialized after years of field trials. Applied agricultural R&D in this new biotechnological era has become more science-based than in the past. It is less dependent on trial-and-error and increasingly dependent upon and interlinked with research in basic biological sciences (Narin, Hamilton, and Olivastro, 1997; Mansfield, 1995). As reported by CHI Research, a typical 1999 biotechnology patent cites nearly twenty scientific publications, an increase even since the mid-1990s and far higher than the one per patent in non-biotechnology fields.

The deepening relationship between science and technology has significant implications for agricultural R&D. Scientists map the broad features of the molecular terrain that technologists explore and partition in a more detailed way, prospecting for new and marketable products. At the same time, scientists themselves make laboratory use of technological innovations and often focus on areas in which technology is advancing rapidly. Communication and coordination between these two is facilitated through the increasingly universal language of molecular and cellular biology. That coordination likely enhances research returns, accounting for much of the rise in agricultural R&D expenditures and of the private sector's share of it.

In America's earlier years, only weak intellectual property protection, such as through the 1930 Plant Patent Act and 1970 Plant Variety Protection Act, was granted to biological inventions (Fuglie et al. 1996). But in last 25 years, patent protection has been granted for an ever-widening array of product and process innovations (Jaffe, 1999). In 1980, the U.S. Supreme Court ruled in *Diamond vs. Chakrabarty* that living organisms were patentable. In 1985, the U.S. Patent Office extended patent protection to all new plant varieties, animal breeds, genes, and traits. These stronger intellectual property rights (IPRs) allow for monopoly profit, enhancing incentives for private biotechnological innovation. Since passage of the Bayh-Dole Act in 1980, which allowed universities to patent innovations resulting from federally funded research, universities have increased their patenting activities and the commercial utilization of academic inventions. Thus, both the private and public sectors have operated in an increasingly privatized setting.

A clear understanding of the rapidly evolving agricultural research system requires an explicit model of the relationship between basic bioscience and applied biotechnology. The competitiveness of U.S. agriculture is determined mainly by its productivity growth. In the new science-based biotechnology, increasing and sustaining agricultural productivity depends not only on the novel technologies developed in applied research, but also on new scientific knowledge generated from basic research and on effective management of the information flows between the two. Therefore, it is important to consider how the public and private sectors are performing in agricultural biotechnology. Does the private sector engage mainly in applied research toward patentable inventions and marketable products, and does the public sector focus on basic research to create the breakthrough scientific knowledge the private sector lacks incentive to conduct? If so, do increasing R&D expenditures and basic bioscience progress enhance a biotech firm's applied research success? Does boosting a university's research budget improve its success in producing basic bioscience? Are basic bioscience and applied biotechnology substitutes or complements for one another? Are agricultural research policies effective in coordinating public and private research efforts and in linking basic with applied research?

To address these issues, I develop below a primal model of agricultural knowledge production. The model distinguishes between inputs and outputs of both basic and applied research in agricultural biotechnology and allows for both complementarity and substitutability between those two research programs. Using data on 1,746 U.S. agbiotech patents issued between January 1985 and August 2000, and the scientific references cited by those patents, I construct a practical measure of information flow between basic bioscience and applied biotechnology. I then design and estimate an econometric model of knowledge output in the basic and applied agricultural sciences and use the model to assess the characteristics of, and relationship between, basic and applied research. Using the econometric results as a framework, I offer guidance to public and private agricultural research policies.

Chapter 2: Literature Review

Since the arrival of endogenous growth theory, the economics of research and development has received increasing interest. Many studies have been conducted to evaluate the returns to public agricultural R&D, employing university research and extension expenditures as shift terms in a production or cost function of individual or aggregated farm outputs (Fuglie, et al., 1996). These approaches seek to link R&D inputs to final-product outputs, permitting direct inferences about the social welfare effects of public research expenditure decisions. With the science-based biotechnology well under way, new approaches are needed, allowing for detailed characterization of agricultural research processes and an explicit examination of the relationships between basic and applied research and between public and private R&D.

However, the agricultural economics literature on biotechnology innovations is still in its infancy, populated mostly by thought pieces and conceptual models. Little explicit or quantitative work has yet emerged on the subject. Econometric tests have been few, and models have specialized on only particular aspects of the innovation and transfer system (Foltz, Barham, and Kim, 2000; Foltz, Kim, and Barham, 2001; Graff, Rausser, and Small, 2001). In general economics, Griliches pioneered the analysis of the relationships between R&D, innovation, and productivity at the firm and industry levels, using patent data from the 1980s. Others developed the patent-based measures further, employing them to examine a number of broadly classified industries. In so doing, they have investigated the production of scientific knowledge, the effects of academic research on industry-patented innovations and productivity change, and both public-private and basic-applied research relationships.

In the first section of this chapter, I discuss the extant literature on the returns to agricultural research. I then: (a) review the literature on private R&D, innovation, and productivity outside of agriculture; (b) summarize studies of public R&D, science, and productivity, and the relationship between basic and applied research; and (c) discuss studies of agricultural biotechnology, particularly those regarding public and private sector R&D, basic and applied research synergies, intellectual property rights, and industry structure.

2.1 Literature on Returns to Agricultural R&D

Sustained use of new knowledge and technology is the cornerstone of American economic growth and development. Economic returns to U.S. public investment in science and technology have been large. The earliest public support for research — over a century ago — was focused on agriculture. Ever since then, research has played an important role in increasing agricultural productivity and enhancing the welfare of agricultural input producers, farmers, consumers, and investors.

Many empirical studies of the social rate of return to agricultural R&D have been conducted. Two main analytic frameworks, the economic surplus approach and the econometric approach, have been used in these studies. The first approach assesses the changes in producer and consumer surplus that may be attributed to research and compares those changes with the associated research cost. Such an approach is usually employed for individual farm commodities. The econometric approach relies instead on statistical estimation of agricultural production processes, in which R&D expenditure is included as an explanatory variable. Using primal or dual methods, research-induced profit gains or cost savings can then be computed, controlling for other factors that may have an effect on agricultural productivity. The latter approach permits one to use a flexible functional form and to assess the returns to agricultural research at a more aggregate level than does the first approach (Fuglie, et al., 1996; Alston, Norton, and Pardey, 1995).

2.1.1 Aggregate Returns to Agricultural R&D

Table 2.1 summarizes studies of the aggregate rate of return to agricultural research investment. Despite differences among these studies in period coverage, data aggregation, model specification, and estimation methods, estimates of annual return rates have consistently hovered between 40 and 60 percent. Some of these studies suggest that the rate of return has declined over time. Fuglie et al. (1996) offer a number of possible explanations for the decline. They argue that a decline might be expected if research expenditures rose relative to the availability of technological opportunities, or if the research funding system had become less effective at selecting the best projects, or if public research had increasingly been directed to nonmarket benefits such as environmental protection and food safety. Yet the evidence for this decline is weak. If we compare measures in recent studies (Huffman and Evenson, 1989; Yee, 1992) with those in earlier ones (Griliches, 1964; Evenson, 1968; Cline 1975), we find return rates to be in the same range in both sets. Even after accounting

Table 2.1. Studies on Aggregate Returns to Agricultural Research and Extension

Author	Methodology	Study Period	Annual Rate of Return (Percent)
Griliches (1964)	Prod. function	1949-59	35-40
Latimer (1964)	Prod. function	1949-59	non-significant
Evenson (1968)	Prod. function	1949-59	47
Cline (1975)	Prod. function	1939-48	41-50
Huffman (1976)	Prod. function	1964	110
Peterson and Fitzharris (1977)	Econ. Surplus	1937-42	50
		1947-52	51
		1957-62	49
		1967-72	34
Lu, Quance, and Liu (1978)	Prod. function	1939-72	25
Knutson & Tweeten (1979)	Prod. function	1949-58	39-47
		1959-68	32-39
		1969-72	28-35
Lu, Cline, and Quance (1979)	Prod. function	1939-48	30.5
		1949-58	27.5
		1959-68	25.5
		1969-72	23.5
Davis (1979)	Prod. function	1949-59	66-100
		1964-74	37
Evenson (1979)	Prod. function	1868-1926	65
White and Havlicek 91979)	Prod. function	1929-72	20
White, Havlicek, and Otto (1979)	Prod. function	1929-41	54.7
		1942-57	48.3
		1958-77	41.7
Davis and Peterson (1981)	Prod. function	1949-74	37-100
White and Havlicek (1982)	Prod. function	1943-77	Jul-36
Lyu, White, and Lu (1984)	Prod. function	1949-81	66
Braha and Tweeten (1986)	Prod. function	1959-82	47
Yee (1992)	Prod. function	1931-85	49-58
Huffman and Evenson (1989)	Prod. function	1950-82	41

Source: Fuglie et al. (1996).

for the deterioration of returns in later years, and the potential upward bias in rate estimates that some critics have pointed out, the annual rate of return to agricultural R&D has apparently been at least 35 percent, much higher than the 18 to 20 percent return rates earned elsewhere in the economy. The high rates of return in agriculture suggest that further allocation of funds to agricultural research would be beneficial to the entire economy. Moreover, the computer- and bio-technologies originating from breakthroughs in the 1950s have not yet been fully exploited in agriculture. With these unexploited technological opportunities, the possibly long lag between research expenditure and payoff, and between scientific discovery and its absorption into technology development, one would expect the returns to agricultural research to remain very high or even to rise in the next 30 to 50 years.

2.1.2 Returns to Agricultural R&D in Crops, Livestock, and Farm Commodities

Besides studies of aggregate rates of return, economists have examined returns to research in the crop sector, the livestock sector, or in particular agricultural commodities. A summary of these studies is provided in table 2.2. Although most research investments have been found to yield high return rates, little consensus has been reached on which agricultural research sectors are the most productive. For example, Bredahl and Peterson (1976), Evenson and Welch (1979), and Norton (1981) found a higher return rate to livestock research than to crop research, whereas Huffman and Evenson (1993) reported the opposite. No strong conclusions seem to be possible about how agricultural research funds should broadly be reallocated to maximize total social benefits.

Table 2.2. Studies on Returns to Agricultural Research and Extension in Crops, Livestock, and Various Farm Commodities

Author	Commodity	Study Period	Annual Rate of Return (Percent)
Griliches (1958)	Hybrid corn	1940-55	35-40
	Hybrid sorghum	1940-57	20
Peterson (1967)	Poultry	1915-60	21-25
Schmitz and Seckler (1970)	Tomato	1958-69	16-46
Bredahl and Peterson (1976)	Poultry	1969	37
	Dairy	1969	43
	Livestock	1969	47
	Cash grains	1969	36
Norton (1981)	Poultry	1969	30-56
	Dairy	1969	27-50
	Livestock	1969	56-111
	Cash grains	1969	31-57
	Dairy	1974	33-62
	Livestock	1974	66-132
	Cash grains	1974	44-85
Sundquist, Cheng, and Norton (1981)	Maize	1977	115
	Wheat	1977	97
	Soybean	1977	118
Smith, Norton, and Havlicek (1983)	Poultry	1978	61
	Dairy	1978	25
	Livestock	1978	22
Huffman and Evenson (1993)	Livestock	1950-82	25.5
	Crops	1950-83	23.5

Source: Fuglie et al. (1996).

The studies discussed heretofore immediately relate agricultural research and extension inputs to final farm production, simplifying social welfare computation. They ignore the intermediary steps in the knowledge generation process, such as how research inputs produce basic research outputs, and how basic research outputs are absorbed by applied researchers and interact with their own research inputs to generate new products, which in turn contribute to agricultural productivity growth.

2.2 Literature in General Economics on Private R&D, Innovations, and Productivity

2.2.1 Private R&D and Patents

Because of the difficulties of obtaining appropriate indicators of research output, empirical analysis of the innovation generation process has been limited. The hypothesis of a systematic relationship between innovation increments and R&D expenditures has been maintained for years. But in the early 1980s, Griliches pioneered the statistical test of this hypothesis in a number of non-agricultural industries, facilitated by the computerization of the U.S. Patent Office's database during the 1970s.

In their 1980 paper, Pakes and Griliches formulated a knowledge production function model in which current innovation increments, measured by the number of patents, are a function of current and lagged research expenditures, a time trend, and a set of firm-specific dummy variables. Using panel data of 121 firms in seven industries from 1968 to 1975, they found a significantly positive relationship between research inputs and outputs. In particular, a 1% increase in current and previous-five-year research expenditures leads to 0.6% increase in patent outputs.

Encouraged by the promising use of patents as an indicator of inventive activity and research output, much work has since emerged on patent-R&D relationships at the firm and industry level. Among them are Griliches (1981, 1984, 1986, and 1988); Hausman, Hall, and Griliches (1984); Pakes (1985); Hall, Griliches, and Hausman (1986); Levin (1987); and Acs and Audretsch (1988). In these studies, the power of R&D expenditures to explain patenting rates was found primarily in cross-sectional observations (cross-firm or cross-industry) rather than in the within-section, time-series dimension.

A major question addressed in these studies was whether there were diminishing returns to R&D. In the cross-sectional dimension, larger firms received fewer patents per R&D dollar, so that in the aggregate, patents-earned did not keep up with the growth of R&D expenditure. But this finding was quite sensitive to functional form, weighting scheme, and the particular point at which the elasticity was evaluated. It has been argued that the appearance of diminishing returns might be due to a selectivity effect in data collection (namely, that small firms are proportionately highly represented in the sample and have to be unusually successful to be listed in the stock exchange and hence included in the researchers' dataset), to the difference between small and large firms in the role of formal R&D, and to the extent to which R&D is reported in small and large firms. In the time-series dimension, the estimated total elasticity of patents with respect to R&D expenditures lay between 0.3 and 0.6;

this estimate was robust to alternative estimation and weighting methods. However, the question of diminishing returns did not arise here, as no strong relationship between annual changes in R&D and patenting is detected in the time series dimension. Finally, patenting culture, such as the propensity to patent, differs greatly across industries.

Various lag structures between research inputs and outputs were tested in the above studies. But the results reveal difficulties in trying to pin down the shape of the lag distribution, and were inconclusive even as to whether a significant lag exists at all (Hall, Griliches, and Hausman 1986). A rather strong contemporaneous relationship between R&D expenditure and patenting was identified, possibly because patents tend to be taken out at an early stage of research and possibly because of some reverse causality: successful research leading to both patent output and to the investment of additional funds for further research and development. A lag-truncation bias due to the possible influence of pre-sample unmeasured R&D expenditures was also found (Hausman, Hall, and Griliches 1984).

Another issue of interest in the R&D-patent relationship is that of R&D spillovers among firms (industries), namely the effect that other firms' (industries') R&D has on the productivity of a given firm's (industry's) R&D. Early evidence of such spillovers, and suggestions for and difficulties of modeling it, can be found in a survey by Griliches (1979). In his 1986 paper, Jaffe quantified some industrial-level spillovers in the R&D process. He first grouped firms into 21 distinct technological clusters, using the distribution of the firms' patents across patent classes. He then constructed a measure of the potential spillover pool in a given cluster, namely the weighted sum of other firms' R&D in that cluster. After accounting for this spillover pool in addition to a firm's R&D and other attributes, Jaffe found that the elasticity of a firm's annual patent numbers with respect to its own R&D expenditures is 0.85. The effect of the spillover pool itself on patent numbers is very large. Firms in clusters in which the average firm performed more R&D received more patents per R&D dollar.

2.2.2 Patents, Productivity, and Stock Market Values

Since the discovery of the "residual", the large portion of output growth that cannot be explained by growth in conventional inputs, in the 1950's, economists have tried to assess the contribution of R&D expenditures to technical change and productivity growth. A popular approach in this effort has been the econometric production function, in which total output or total factor productivity is a function of conventional inputs and R&D expenditures. As discussed in section 2.1, the same approach has also been widely used in agricultural economics. A rich literature is available on this topic, including Griliches (1964, 1979, 1986, 1988, 1994), Terleckyj (1974), Mansfield (1980), Lichtenberg and Siegel (1991), and many others. The main finding of these studies has been a significantly positive relationship between various measures of productivity and R&D expenditures at the firm, industry, or country level. But estimated R&D effects are modest, not large enough to account for observed productivity fluctuations.

With the help of computerized patent data, a number of studies have attempted to link patent numbers with measured productivity growth, assuming patent numbers are indicators of inventive outputs, which in turn contribute to productivity (Basberg 1982; Walsh 1984; Eaton and Kortum, 1996; Porter and Stern, 2000). The estimated relationships have been inconclusive and Griliches offers possible explanations for this in his 1990 survey. He reasons that: (i) at most only one half of productivity growth is due to innovation increments, and only a fraction of this is captured by patented inventions; and (ii) an invention's effect on productivity likely requires long and variable lags, which cannot be identified with the available data and are further smoothed by aggregation over many inventions.

As the success of a firm's research effort is thought to be reflected in its stock market value much more quickly than in its profit or productivity, another line of research (Griliches 1981; Pakes 1985; Jaffe 1986; Cockburn and Griliches 1988; Griliches, Hall, and Pakes 1991; Hall 1993; Blundell, Griffith, and Reenen 1999) has investigated the value of patents, using data on firm market values. By estimating a firm-level market value function, including physical tangible capital and intangible capital reflected in patents or R&D, some researchers have found that patents typically do not have as much market-value explanatory power as does R&D, even though patents should be suggestive of the "success" of a firm's R&D program. The low correlation between patent rates and dollar-denominated measures such as R&D and market value might arise because patent counts are a noisy measure of the economic significance of innovations contained in these patents. The distribution of patent qualities is known to be dispersed and highly skewed. Few patents are very valuable, and many earn no revenues at all. The number of patents a firm holds cannot, therefore, well represent the sum of values of those patents, and we should not expect the correlation to be high.

2.2.3 Patent Values and Other Uses of Patent Data

With more information on patent documents becoming available since the late 1980's in computerized form, some economists have focused on improving the conceptualization and measurement of research outputs. The ideal measure of a patent's quality would be the license revenues eventually earned from it. Unfortunately, data on license revenues are hard to find. Alternative measures, such as the number of claims in a patent, the patent's renewal, and the number of countries in which an invention is patented, have been used as proxies for patent quality (Lanjouw and Schankerman, 1997). Among these measures, patent citations have attracted the most interest. Patent citations represent the previously existing technological state-of-art upon which a citing patent builds, so that the number of citations a patent receives provides evidence of its technological importance or even economic value. Using patent citation data, Trajtenberg, Henderson, and Jaffe (1997) constructed several measures of "importance" and "generality" to capture the "basicness" and "appliedness" of a matched sample of university and corporate patents. They suggested reexamining older studies in the light of their measures instead of using simple patent counts as R&D output indicators. Several analysts have explored the relationship between citation-weighted patent rates and value measures such as licensing revenues (obtained from a limited sample survey), stock market values, and social welfare (Trajtenberg 1990; Harhoff et al.1999; Hall, Jaffe, and

Trajtenberg 2000). They find that those value measures are considerably more highly correlated with citation-weighted patents than with the simple patent counts, reconfirming that citation weighting provides a superior proxy of patent quality.

Another important signal that patent citations convey is knowledge flow, which otherwise might have been regarded as nearly impossible to trace. A large amount of noise is recognized to exist in citation data (Jaffe, Trajtenberg, and Fogarty, 2000). For example, some patents' citations to previous patents are added by the U.S. patent examiner, so that the citing-patent inventors were unaware of the cited patents during the invention process. In this case, it is clear that no direct knowledge spillovers have occurred. However, the "objectivity" of such citations may be greater than those added by the inventor, and they may provide important contributions to the technological context in which the patent was granted (Griliches, 1990). In that sense, even the patents entering the citation list through the patent officer have spilled into the citing patent, though not in an explicit manner. As patents contain detailed information about application and issue date, patent class, assignee organization, and geographic location of inventors, economists can use patent citation data to investigate knowledge flows across time, technological, organizational, and space boundaries. For example, by comparing the geographic location of the citing patent with that of the cited patent, Jaffe, Trajtenberg, and Henderson (1993) find that the knowledge spillovers are geographically localized at country, state, and local (SMSA) levels, although the localization fades slowly over time. They find no evidence of a technological area effect on citation localization, that is, citations in the same patent class are as likely to be from the same geographic location as not. They attribute this, as did Jaffe (1986), partly to the fact that knowledge spillovers are not confined to a firm's immediate technological neighborhood.

The National Bureau of Economic Research has developed a patent database comprising detailed information on U.S. patents issued between 1963 and 1999 and all citations made to these patents between 1975 and 1999. Documentation and description of this data file can be found in Hall, Jaffe, and Trajtenberg (2001). Some variables in the file are based on patents' front page information, such as technological category, citation lags, number of citations made and received, share of self-citation (citation to patents owned by the same organization as the citing patent), and measures of generality and originality. Hall, Jaffe, and Trajtenberg also matched their patents to firms listed in 1989 in Compustat (the production and financial data of all firms traded on the U.S. stock market), hoping to use the matched data to examine the relationship between R&D, patents, and firm-level production. It is the most comprehensive public patent and citation data file thus far developed.

2.3 Literature in General Economics on Public R&D, Science, and Productivity

Work in the economics of science has focused mainly on three lines of inquiry (Stephan 1996). Inspired by earlier work of sociologists on incentive schemes for scientists and scientists' reactions to those schemes, economists have demonstrated the existence of a non-market-based reward structure in science and have examined the characteristics and efficiency of this structure (Merton 1957; Dasgupta and Maskin 1987; Dasgupta and David 1994; Stern 1999). In a second line of inquiry, analysts have sought to understand the scientific labor market in a human-capital framework, enabling an explanation of scientists' publishing activities, acceptance of new ideas, and earnings over a life cycle. Studies along this line include Becker (1962), Levin and Stephan (1991), Ehrenberg (1992), and Leslie and Oaxaca (1993). As the most important factor attracting economists' attention to science is the indisputable contribution of science to economic growth, the third line of inquiry is about the production of scientific knowledge itself and about the relationship between science and technology, for example, about how each spills over to facilitate the other. Studies of the effects of technology and innovation on profitability and growth have been discussed in the previous section. In the following, this third line of inquiry about the economics of science is reviewed.

2.3.1 Public R&D and Scientific Papers

On the production of scientific discovery and knowledge, economists traditionally have focused most of their attention on the direct contributions of an individual scientist's time and cognitive resources, such as his intelligence and knowledge base (Stephan 1996). Although, in his 1959 paper, Nelson recognized the risky nature of basic scientific research and the substantial economic resources required to pursue it, economists didn't begin quantifying the role that research resources play in knowledge production until about two decades ago.

Parallel to the framework discussed above in the analysis of patents and R&D expenditures, knowledge production functions are often used to explore basic research input-output relationships, both in agricultural science and in other scientific fields. In these knowledge production functions, knowledge increments depend on current and lagged research expenditures, a time trend, and other variables. Scientific publications, or the subsequent citation performance of these publications, usually are used to proxy the quantity (or quality-adjusted quantity) of scientific output. However, this topic has been far less studied than has the relationship between industrial patents and R&D.

Using a panel of publication outputs and research expenditures in the 48 mainland agricultural experiment stations, Pardey (1989) quantified the input-output relationship in agricultural research. He found no systematic effect of year-to-year, within-state research expenditure fluctuations on research performance, while between-state differences in average research expenditures did show a systematic influence on scientific publication rates. This is consistent with the finding from non-agricultural patent and R&D studies that the explanatory power of R&D expenditures exists primarily in cross-sectional observations rather than in the time-series dimension. Pardey's results convey little information about the precise shape of the lag distribution. However, his study does obtain a significant relationship between long-run lagged research expenditures and publication output, as measured at least by the sum of coefficients on current and lagged expenditures. He estimates that a 1% increase in total research expenditure during the previous seven years led to a 1.2% increase in publication output. Using citation performance to adjust the quality of publication output, this expenditure response rises to around 1.6%. Alternatively, using quality-unadjusted and quality-adjusted output measures, mean gestation lags between project inception and completion were 2.8 and 3.4 years, respectively.

In their 1996 paper, Adams and Griliches focused on the research performance of U.S. universities in eight scientific fields during the 1980's. In most of scientific fields aggregated across universities, they found approximate equality between growth rates in publication and citation and that in R&D expenditures, suggesting constant returns to scale in the production of new scientific knowledge. Exceptions were in agricultural science and mathematics, where increasing and decreasing returns prevailed, respectively. At the university-field level, diminishing returns to research scale were found in every field. The elasticity of research output with respect to research input was higher using citations rather than publications as the output measure. In agricultural science, a 1% increase in R&D expenditures brought a 0.90% and 0.93% increase in publication and citation quantity, respectively. In all eight fields taken together, the average elasticity of publication and citation quantity with respect to R&D expenditures was 0.60 and 0.73, respectively.

The contrast between constant returns at the aggregated field level and diminishing returns at the university-field level might be explained by the exclusion of research spillovers between universities at the university-field level and possible misclassification and imperfect accounting for R&D by university and field. Adams and Griliches conclude that the leading schools are more efficient in generating scientific publications than are lesser schools, and private schools are more efficient than their public counterparts. An equation for graduate teaching output, measured by the number of Ph.D. degrees awarded in each field, was added to the research output equation to form an equation system. In SUR estimates of this system, R&D impacts on publication and citation quantity were similar to those in the single-equation estimates. The number of Ph.D. degrees was significantly positively associated with universities' quality rankings.

2.3.2 Scientific Papers and Productivity

As applied R&D has not been observed to be a large contributor to productivity growth, several economists have argued that something more basic, i.e. the expansion of the knowledge base opening up new technological opportunities, has been left out of the modeling process. Adams (1990) has sought to overcome this discrepancy by exploring the relationship between published academic science and multifactor productivity growth in 18 manufacturing industries between 1953 and 1980. He developed an indicator of the stock of knowledge in each of nine scientific fields at a given time by counting the number of publications in that field from 1930 forward. An industry-level knowledge stock measure was then created by allocating these counts according to the distribution of scientific personnel in each industry field, under the assumption that an industry's ability to absorb advances in basic scientific research is maintained by the scientists working in that industry.

Adams also distinguished between an industry's own knowledge stock and the "spillover" knowledge stocks absorbed from other industries, weighted by the technological closeness of those industries to the industry in question. By regressing annual productivity growth in the eighteen industries against own and "spillover" knowledge stocks, energy shock variables, and industry dummies, Adams found both knowledge stocks to be sizable contributors to productivity growth. The elasticity contribution of own knowledge to productivity growth ranged between 0.30% and 0.86%, and of spillover knowledge to productivity growth between 0.11% and 0.64%, in various specifications and subperiods. Adams identified a lag of about 20 years for own knowledge and 30 years for interindustry knowledge spillovers, implying a much longer search time than those identified in the private R&D studies discussed above.

2.3.3 Public and Private R&D and Basic and Applied Research

2.3.3.1 Universities' Basic Research and Firms' Applied Research

Basic research directed toward the advancement of scientific knowledge, instead of toward the creation of new or improved products and processes, involves a large amount of resources and great uncertainty, and returns to it are largely inappropriable. Hence, a gap exists between the marginal private and marginal social benefits from this research, and private firms lack incentives to perform socially optimal levels of basic research (Nelson 1959). Universities traditionally have been dedicated to basic research and to the free dissemination of its results. Despite recent trends toward the privatization of their research, most university scientists still focus on relatively abstract concepts which have no immediate application to goods or services. Instead, they are responsible for invoking scientific breakthroughs, opening up new technological opportunities, and reducing the costs of applied research. Therefore, most university research has economic value only insofar as it affects the subsequent development, primarily by private firms, of commercially viable inventions and products. This is often called the spillover effect of university research.

Some economists have worked to characterize the relevance of university research and teaching to technical innovations, or to their scientific antecedents, at commercial firms. For example, in surveys of research managers in 130 industries, Nelson (1986) asked respondents to score, on a scale from 1 to 7, the relevance in their lines of business of various university research and teaching fields to technical change. He found that university research was an important source of technical progress in certain industries, particularly in those related to the biological sciences. University teaching on average received higher score than university research in that it provides training to future industrial scientists.

Mansfield (1991) surveyed 76 firms in seven manufacturing industries to seek the scientific roots of their product and process innovations commercialized from 1975 to 1985. He found that 11% of their new products and 9% of their new processes could not have been developed (without substantial delay) in the absence of academic research carried out during the 15 years preceding the commercialization. The mean lag between the relevant academic research finding and the product's or process's commercial introduction was about seven years. Using sales data, he estimated the social rate of return to academic research to be 28 percent. Mansfield (1995) asked each of the 76 firms to cite academic researchers whose work contributed most importantly to their new products and processes. He then linked those researchers to data on the resources and characteristics of the universities at which they worked. He found that the contribution of a university's research to industrial innovation was related directly to the quality of its faculty, to the size of its R&D expenditures in relevant fields, and to its geographical proximity to the industries affected. The academic researchers cited by these firms reported complementarity between their government-funded and industry-funded work: their government-funded research was more fundamental and preceded their industry-funded research, and the ideas and problems they encountered in their industrial consulting influenced their government-funded projects.

Although these surveys provide evidence of spillovers from universities to firms, Jaffe (1989) was the first to identify the extent to which university research relates to the generation of private-firm innovations. In order to model this relationship, he modified the knowledge production function introduced by Pakes and Griliches (1980). In addition to industry R&D expenditures, university research expenditures and a measure of the geographical coincidence of university and corporate research were included as inputs to produce firms' new innovations as measured by patent counts. To allow for endogenous determination of university and industry R&D expenditures, the knowledge production equation and two other equations characterizing the interdependence between university research expenditures and industry R&D were estimated simultaneously, using state-level time-series data in five broad technology areas.

Jaffe's results provide evidence that corporate patenting activity responds positively to commercial spillovers from university research conducted in the same state. The elasticity of response is about 0.1. University research appears to have an indirect effect on local innovations by inducing commercial R&D within the state. Combining the elasticity of commercial R&D with respect to university research (0.704) and the elasticity of corporate patents with respect to commercial R&D (0.814) gives an elasticity of induced corporate patents with respect to university research of about 0.6. On the other hand, the effect of commercial R&D on university research expenditures is fairly small and nonsignificant. Using a more direct measure of innovative output, the number of innovations instead of the patent count measure in a knowledge production function, Acs, Audretsch, and Feldman's (1992) estimation results reinforced Jaffe's (1989) findings. There were some indications that knowledge spillovers, particularly those from university research, are more decisive in promoting innovation activity for small firms than for large firms (Acs, Audretsch, and Feldman, 1994).

Geographical localization of knowledge spillovers is embodied in these studies to the extent that spillovers can flow only within a state boundary. Indeed, Audretsch and Feldman (1996) find that knowledge-based industries tend to cluster their innovative activity and production near areas of university research and skilled labor. The rationale for this localization is that part of the knowledge remains tacit in the sense that it is difficult to communicate in writing, but instead is facilitated through personal communication and cooperation. Yet the actual mechanisms by which the knowledge is transported have not been modeled. Without a trail linking knowledge-producing universities with the firms using their knowledge, it is difficult to ascertain whether knowledge spillovers are truly geographically bounded. A more structural model of the connections between academic science, private R&D, and patenting rates, estimated with data on individual universities and firms, would lead to better understanding of the mechanism underlying the nation's research system.

2.3.3.2 Universities' Applied Research, Firms' Basic Research, and their Implications for the Basic-Applied Research Relationship

I have focused above on the public-good characteristics of universities' basic scientific research, and on the proprietary nature of private firms' technological innovations. However, since the 1980 Bayh-Dole Act and subsequent court decisions permitting universities to patent many of their own innovations, the university sector has operated in an increasingly privatized setting. Universities have established technology transfer offices to identify patentable inventions, to market these inventions by licensing their use rights to large private or start-up companies, and to seek industry support for university research (Graff, Heiman, and Zilberman 2001). From 1985 through 2000, university patent awards increased from 550 to 3,272 per year, while the share of university patents in total U.S. patents rose from 0.5% to 2.2%. Universities have increasingly commercialized their research results: in 2000, university licensing revenues amounted to $1,100 million, up from $130 million in 1991, and the start-up companies founded on the basis of university research numbered 368, up from 223 in 1995 (AUTM Licensing Survey, various years). As noted by Zilberman, Yarkin, and Heiman (1999 and 2001), most of the rise in university patenting and licensing activities has occurred in the medical and biological fields.

Henderson, Jaffe, and Trajtenberg (1998) offer a detailed analysis of 1965-1992 university patenting. They find, using citation-based measures of importance and generality, that university patents prior to the mid-1980's were more important and more general than were a random sample of all U.S. patents. Since then, this advantage has disappeared. Legal changes such as the Bayh-Dole Act apparently have increased universities' propensity to patent and to commercialize, but have had no significant impact on the generation of commercially important university inventions.

In a more recent study, Thursby and Thursby (2000) have modeled university licensing as a three-stage process, each involving multiple inputs, to identify the sources of growth in university licensing. Using data from an AUTM licensing survey and from a survey of firms which licensed university inventions, they find results similar to those of Henderson, Jaffe, and Trajtenberg's (1998), that is, that increased licensing has been due more to an increased willingness of faculty and administrators to patent and to license than to a shift in the faculty's research agenda.

Privatization of university research has been reflected also in the rise in industry support for university research, which may in turn be a response to universities' rising interest in applied research and patenting. The last two decades have witnessed a dramatic increase in public-private research partnerships. These have drawn substantial attention from economists, most recently from Hall (2002), Adams, Chiang, and Jensen (2000), Adams, Chiang, and Starkey (2000), Hall, Link, and Scott (2000), and David and Hall (2000). While some worry that university researchers have become too responsive to economic incentives and that public-private partnerships might crowd out public-good university research, others believe that privatization fosters knowledge spillovers and that the commercialization of university research helps realize more complementarities between basic and applied research and between public and private R&D (Rausser 1999).

Just as universities emphasizing basic research still do applied research, private firms engaged primarily in applied research and development also do basic research. Cohen and Levinthal (1989) recognized the dual role of a commercial firms' R&D: innovation and learning. By conducting basic research, firms develop their "learning" or "absorptive" capacity to evaluate and exploit potentially useful scientific knowledge created in universities or in other sectors. They thereby gain a first-mover advantage in exploiting these new technologies. Alternatively, they are able to assimilate new technologies that competitors have developed and thus act as a rapid second mover in the presence of inter-firm spillovers. Mansfield (1980) was one of the early to relate an industry's or firm's productivity change to the amount of basic research it has performed. With the help of survey data on the R&D expenditure composition of 119 firms, he found a statistically significant and direct relationship between basic research effort and total factor productivity growth at both the industry and firm level, holding constant applied research effort. This finding suggests that firms which direct their resources more toward basic research have a larger capacity to utilize information available elsewhere, or that applied R&D is more effective when carried out in conjunction with basic research.

In recent years, applied research has become increasingly science-based, less dependent on trial-and-error. As Arora and Gambardella (1994b) point out, increases in theoretical understanding, better instrumentation, and better computing capabilities have universalized the terms in which applied R&D is conducted, fostering greater communication between scientists and applied technologists. The increasing universality of technology has enabled a greater division and specialization of firms' research efforts and has brought greater pressure for firms to seek complementarities between their basic and applied research (David, Mowery, and Steinmueller 1992). Cockburn, Henderson, and Stern (1999) provide an example of this complementarity in the design of commercial firms' incentive schemes for research effort. A "basic" research management strategy is to base an individual employee's promotion and salary on his publications, whereas an "applied" strategy is to reward an entire group's patenting success by increasing the work budget for that group. These two strategies are complements in the sense that raising incentives in one direction increases the marginal return to raising incentives in the other. Cockburn, Henderson, and Stern show that weighting publications strongly in determining an employee's compensation tends also to reward her patenting success highly, exploiting the complementarity between scientific and technological skill.

In our rapidly evolving research system, the public and private sector each conducts both basic and applied research, scientific and technological knowledge spill between the two sectors, and public and private funding both interact and compete with one another. Hence, the system is multi-dimensional and complicated. Economists can significantly contribute to an understanding of this system and can guide research policy design if they employ appropriate measures of research inputs and outputs and develop accurate structural models of research processes and interactions.

2.4 Literature on Biotech and Agricultural Biotech R&D

As the agricultural biotech industry is still in its infancy, the agricultural economics literature on biotechnology has contained little quantitative analysis and instead has proceeded in exploratory fashion. Studies in agbiotech range between discussions of policy issues, organizational descriptions of the industry, thought pieces on international development, rough estimates of surplus, and conceptual models of research, production, and distribution. Because of the complexity in examining information flows and the difficulty in modeling the relationships between public and private and basic and applied research, especially few studies of the characteristics and productivity of agbiotech R&D have emerged.

2.4.1 R&D in General Biotech

Fortunately, one can learn from studies of R&D in general biotechnology, or in particular, in pharmaceutical biotech, the more mature cousin of agbiotech. Researchers have employed knowledge production functions to explore the relationship between industrial R&D, patents, productivity, and stock market valuations in the chemical, biomedical, and pharmaceutical industries, among others, as discussed in section 2.2.

2.4.1.1 University Bioscience Research and Commercial Biotechnology Research

The heavy dependence of the biotech industry on basic science research has led a number of researchers to examine the role of geographic proximity between the biotech firm and the university scientist. Audretsch and Stephan (1996) collected data from 54 biotech firms that had prepared an initial public offering in the U.S. between March 1990 and November 1992, identified 445 university-based scientists affiliated with those firms and the role each played with the firm, and linked the scientists to the name and location of their home universities. By doing so, they developed an explicit paper trail linking biotech firms and university scientists.

Their data suggest that although biotech firms were geographically concentrated, the supply of scientific talent was much less so. While locational proximity plays a role in establishing ties between firms and scientists, it is by no means an overwhelming role. Approximately 30% of university scientists affiliate with firms in the same region they are, but the rest of the firm-scientist links are not geographically bounded. Audretsch and Stephan identified three key roles that university scientists played in biotech firms: they facilitate knowledge transfer, signal the firms' quality to the scientific and financial community, and help chart the companies' scientific direction. They found that scientists who performed the first function were more likely to be local than were scientists performing the latter two functions. Scientists' characteristics, such as their age, citation history, and Nobel Prize status, also shaped the importance of geographical proximity to the scientist-firm connection. They noted that the locational link was more important when knowledge spillovers were informal. When knowledge was transmitted in a formal setting, geographical proximity did seem less valuable.

Zucker, Darby, and Brewer (1998) believe that, at least for the first 10 to 15 years of their history, biotechnology innovations have been based on naturally excludable knowledge held by only a few "star" scientists. To test the hypothesis that the entry of new biotech firms is explained primarily by the distribution of leading scientists who actively contribute to basic science, they used panel data (in 183 functional economic areas from 1976 to 1989) on new entrants and incumbents in the biotech industry. The data consist of "star" scientists identified through publications in which gene sequence discoveries were reported in academic journals, the organizational affiliations of these "star" scientists, information on the associated regional scientific base as reflected in the "biotech-relevant" university department ranking and the value of federal research support, and local economic activity indicators.

Regressing the number of biotech firms per area against the number of "star" scientists in that area and against a set of scientific base and economic activity indicators, they find that the quantity of important scientific findings was the principal determinant of the growth and location of the biotech industry. Research universities and their highly productive scientists played a key role in this new high-tech industry above and separate from the federal research funding they received. Thus, biotech start-up firms were more likely to be found in locales with outstanding university bioscientists, probably because geographic proximity reduced the cost of exploiting the tacit and complex scientific knowledge required to produce an innovation. The Zucker-Darby-Brewer study, however, lacks a paper trail of information flows between scientists and firms. It can demonstrate localized effects but cannot affirm or deny the presence of knowledge spillovers in a local knowledge network.

2.4.1.2 Commercial Biotech Research and University-Firm Collaboration

Because of the complexity of the knowledge base which underlies biotechnology, and because of biotech's science-push origin, biotech firms' absorptive capacity for external information has become increasingly important. So too have the collaborative alliances among large biotech firms, small new firms, and universities. Consistent with Cohen and Levin (1989), Arora and Gambardella (1994a) distinguish between two types of biotech firm capabilities: the scientific ability to evaluate information, and the technological ability to use it. In order to characterize how these two abilities help a firm to enter into and benefit from collaborative relationships, Arora and Gambardella compare large biotech firms' collaborative agreements with their in-house science bases and technological capabilities, respectively measured by their published scientific papers and biotech patents, and with a set of firm characteristic variables. Using data on a sample of 26 established large U.S. biotech companies, they concluded that a firm's technological ability increased the number of its external linkages, while its scientific ability made it focus on fewer but more valuable collaborations. They acknowledged they could not capture much of the complexity of technical and knowledge linkages.

Another indicator of a firm's tacit knowledge capture, the number of research articles coauthored by firm scientists and top research university scientists, was developed by Zucker, Darby, and Armstrong (2002). They hypothesized that collaborative work with university scientists enables commercial firms to capture the scientists' tacit knowledge, which in turn promotes the firms' innovation and product development. Using panel analysis, they find that publications written jointly by firm and university scientists have a positive effect on the number and citation performance of the firm's patents, employment, and output. Papers coauthored with "star" scientist are more significant in this respect than are other joint papers. Although the Zucker-Darby-Armstrong study explicitly traces the paper trails between biotech firms' and universities' basic research, it does not document the information flows from basic to applied research. Given that the basic research represented by the joint journal articles presumably feeds into the firm's applied research, the contribution of basic research from sources other than the coauthored papers appears to be neglected.

2.4.2 R&D in Agricultural Biotech

2.4.2.1 Intellectual Property Rights, Industry Structure, and Agbiotech R&D

Strengthening intellectual property protection and the changing structure of the biotech industry have reshaped agricultural research, requiring an analysis that goes beyond the conventional agricultural R&D models reviewed in section 2.1

(Shoemaker et al. 2001). Yet, just as the agbiotech industry itself, the analysis of agricultural biotech R&D clearly is an infant industry, marked largely by conceptual models and descriptive analysis.

Most agricultural research outputs historically were released into the public domain or held as trade secrets, since property-right protections for them were difficult to enforce because of their public-good characteristics, that is their nonexcludability or nonappropriability. But the arrival of biotechnology redefined intellectual property. Since the 1980 Supreme Court's *Diamond vs. Chakrabarty* decision allowing patents on genetically modified living organisms, and the Bayh-Dole Act permitting patents on discoveries from federally funded research, patenting opportunities have expanded for public and private agbiotech research. In response to the potential monopoly profits which patents confer, perhaps the most powerful of all intellectual property protections, private companies have been actively and increasingly engaged in agricultural research. In 1998, 57% of agricultural R&D in the United States was performed in the private sector, an increase of about 22% from 1980 (USDA). In contrast, public R&D funding for agricultural research remained almost constant in real terms during the same period. The public sector increasingly has sought patent protection for its agricultural research, and the revenues from the commercial utilization of these patents. For example, Barham, Foltz, and Kim (2001) report that university agbiotech patent volume rose from 10 per year in the early 1980's to 105, 124, and 174 in 1997, 1998, and 1999, respectively. University agbiotech research has in some sense been privatized through this patenting trend.

Models have been developed of the relationship between intellectual property rights and agbiotech R&D. Recognizing that an agent inventing a new agricultural technology would, when transferring the technologies to a competitive farm sector, exploit the monopoly rights afforded by its patent, Moschini and Lapan (1997) constructed a profit function to evaluate the implications of intellectual property rights for the welfare effects of agricultural R&D. Using simulated data, they find that conventional models which assume competitive pricing of publicly produced innovations usually overestimate the welfare gains from agricultural R&D. In their 2000 paper, Oehmke et al. focused on how increasing private ownership of intellectual property affect the agribusiness environment and the evolving role of public agricultural research. A dual cost neo-Schumpetarian framework, in which R&D is modeled as a sequence of stochastic races, was employed to examine whether the commercialization of public research maximizes social welfare. Oehmke et al.'s model includes two major types of life-science company: large firms with well-funded R&D, and small university-related firms. Their results suggest that, relative to the social optimum, private firms underinvest in applied research. But there is a role for the public sector in conducting R&D in niche markets and in providing the basic research which enhances the productivity of applied research.

Koo and Wright (1999) utilize a stylized dual model of cumulative innovation to explore the dynamics of patent protection with licensing agreements for plant genetic resources. They find that the patent life and royalty rate which maximize worldwide dynamic social welfare would differ from those which optimize static social welfare.

Intensive mergers, acquisitions, and co-venture activity in agriculture biotechnology during the last decade have dramatically increased concentration in this industry. In 1997, the leading four-firm concentration ratio in the U.S. corn, cotton, and soybean seed market reached 67%, 71%, and 49%, respectively (Hayenga 1998). Some researchers (Kalaitzandonakes and Hayenga 1999) have argued that these horizontal and vertical integrations have been driven by the characteristics of biotech research itself, by the strengthened patent system, and by high transaction costs. Consolidated firms with a large market share and strong patent portfolios are in a better position to exploit complementary knowledge assets and thus to capture greater returns to R&D. Too much market power, however, might reduce R&D activity in this industry. Concentrated firms have greater access to capital and to agbiotech markets than do start-up firms, which often conduct the research bridging the gap between basic research and new product development. Moreover, concentrated firms' strong patent portfolios may inhibit public and private research by limiting access to newly patented technologies. Brennan, Pray, and Courtmanche (1999) investigated the relationship between agbiotech industry concentration and R&D activity, the former measured by the four-firm concentration ratio or Herfindahl-Hirshfield Index, and the latter by the quantity of field trials. Using an innovation market approach, they showed that concentration had not reduced R&D activity as a whole, but had adversely impacted the R&D activity of non-top-four firms and inhibited the entry of new firms. There was some evidence of reduced research efficiency among large firms.

Like pharmaceutical biotech, agricultural biotechnology is science-based. But biotechnological innovations are used extensively by bioscientists, and successes and failures in new technology development affect the topics in which bioscientists are involved. Therefore, as Rausser (1999) has discussed at length, applied research efforts can enhance basic research just as basic research insights facilitate technological development. That is, a complementarity exists between basic bioscience and applied agricultural biotechnology. How to encourage and exploit such complementarities is of great interest to R&D policy makers, research administrators, and private firms. Yet no explicit study addressing these issues has emerged.

2.4.2.2 Patent Production in Agbiotech

Patent production in agricultural biotechnology in been investigated in several empirical quantitative studies. Foltz, Barham, and Kim (2000) focus on university patent production. They identify 795 agbiotech patents owned by 107 universities, with application dates ranging from 1971 through 1998. Following the knowledge production function approach, they relate each university's agbiotech patent numbers and citation-weighted patents to its research funding by source (federal, state, industry, and own institution), to the quality of its research labor as measured by graduate school and biology-related department ranking, to its land grant status, to the number of agricultural science Ph.D.s awarded, and to the number of employees in its

Technology Transfer Office, among other variables. Their results reveal the importance of land grant university infrastructure, technology transfer effort, and the presence of "star" scientists in a university's agbiotech patent production. While own-institution research expenditures are potentially important, funding from federal and state governments and industry do not in their study seem significantly to promote patent-producing research.

This static analysis was extended by Foltz, Kim, and Barham (2001) to capture the dynamic and nonlinear processes of university agbiotech innovations. They developed a dynamic count-data model in which a feedback mechanism from previous patent success is incorporated explicitly into the modeling structure. In addition to the factors found important in the static analysis, the dynamic model demonstrates that state funding and feedback affect patent success as well.

Graff, Rausser, and Small (2001) investigate private-firm production of three technological categories of plant biotechnology patents, test the complementarities between these intellectual assets, and examine the relationship between changes in industry structure and intellectual asset complementarities. Data were collected on 1,188 agricultural biotech patents awarded to 76 firms between 1975 and 1998, on the firms' R&D spending, and on their other characteristics such as their total sales, share of all plant biotech patents, and non-plant-biotech patents. Graff, Rausser, and Small sort patents into three key categories of plant biotechnology: genetic transformation tools, gene sequences and genetically coded traits, and elite plant germplasm. They aggregate the data on those 76 firms into consolidated and unconsolidated industry samples based on the industry's pre-merger and post-merger structure in 1994 and 1999, respectively.

Regressing patent counts in each of the three categories against firm R&D expenditures and other characteristics, they find in all three patent categories that regression fitness was much higher whereas the coefficients significance was lower in the consolidated industry sample than in the unconsolidated one. As expected, a firm's share of total plant biotech patents had a significantly positive effect on its patent production, while its size, approximated by its sales and its non-plant-biotech patent production, had positive but insignificant effects. Complementarity between two classes of knowledge stock is indicated by a positive correlation in the residuals of the patent production regression of those two classes. The strongest complementarity was found between genetic transformation patents and gene-sequence patents. Some complementarity was detected between genetic transformation and germplasm patents, and between gene-sequence and germplasm patents. Patent complementarities rose when moving from the unconsolidated to the consolidated industry sample, suggesting that the recent dramatic consolidation in the biotech sector can partly be explained by the desire to exploit these complementarities.

All the studies discussed in this section have concentrated on either a university's or private firm's biotechnological research. Thus they provided little information about the characteristics of basic bioscience research or about the relationships between basic bioscience and applied agricultural biotechnology research.

The growing interdependence between science and technology in agbiotech, the increasing privatization of agricultural research in both the public and private sectors, and rising concentration in the agbiotech industry have profound implications for public and private agbiotech research activities and for the relationship between basic bioscience and applied biotechnology in each sector. Thus, a clear understanding of the returns to agricultural research in this rapidly evolving biotech era warrants a more detailed and explicit study of the relationship between basic and applied research in the public and private sectors.

Chapter 3: Theoretical Framework

To assess the efficacy of, and relationship between, basic and applied research in public and private sectors, a theoretical model is developed here distinguishing between inputs and outputs of both basic and applied research and permitting both complementarity and substitutability among research programs.

3.1 Identifying Information Flows

In the human-capital-intensive U.S. economy, and especially in biotechnology, productivity is enhanced largely through a reorganization and intensification of knowledge. The study of the efficacy of basic bioscience and applied biotechnology research conducted by public and private research enterprises, or in Arora and Gambardella's (1994b) words, the "technology of technical change," essentially is the study of information transformation, that is, of how the information inputs are transformed into information outputs. Thus, increasing and sustaining agricultural productivity in the biotech era depends greatly on the management of information flows among bioscientists and biotechnologists (Zilberman, Yarkin, and Heiman 1999).

Unfortunately, information variables cannot be traced as easily as physical inputs are. It would be difficult even for a researcher herself to identify all the information leading her to a particular insight or discovery. Yet because ideas are fruitful only in combination with related ones, they are best bundled in such printed forms as patent documents, books, journal articles, and theses and dissertations, or in unprinted form such as meeting presentations and online publications. This sort of bundling provides us with the possibility of identifying some of the major sources of inspiration and prior art of a research project.

However, document bundling has its limits. Scientists cannot put in writing everything they know about a genetic transformation tool or gene sequence they are working on. Part of their knowledge has to remain "tacit" in the sense that practical information about it can be provided only through continuous face-to-face communication. This was especially true during the early stage of biotechnology development. As discussed in section 2.4.1, some economists have shown that the tacit portion of knowledge explains why start-up biotech companies tend to locate close to research universities with strong biology research programs and "star" bioscientists. It also explains why biotech companies employ basic scientists and form cooperative alliances with universities, since only by doing so can they enhance their "absorptive capacity" to evaluate and capture the knowledge sources available.

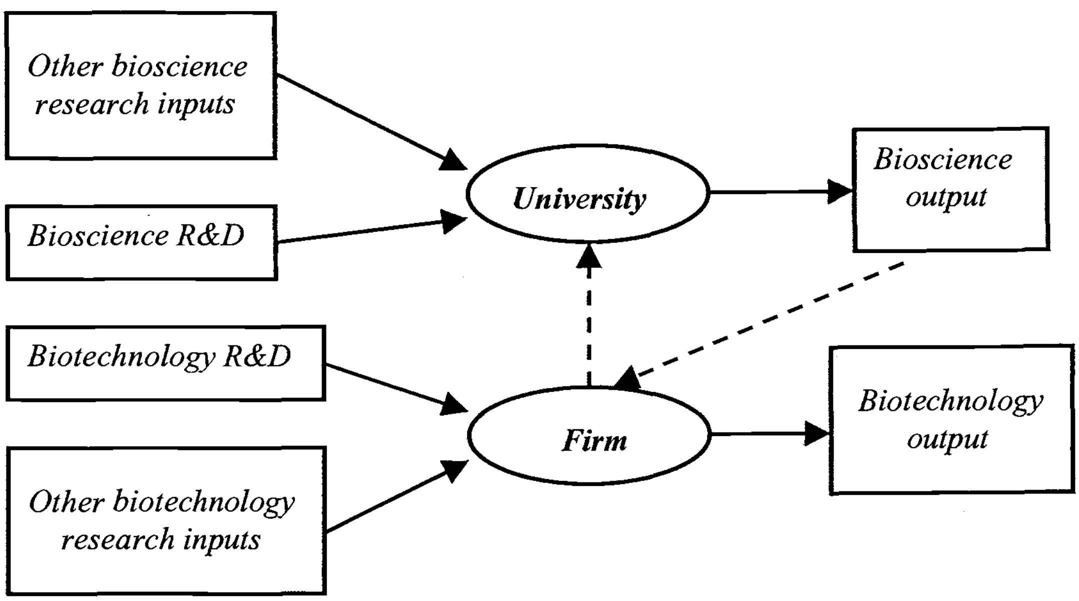
In the present study, I am interested in the information transformation between basic bioscience and applied biotechnology research. As one can imagine, it is difficult to draw a clear line in practice between what is "basic" and what is "applied." Although in principle it is widely agreed that basic research is directed toward an increase of fundamental knowledge and all understanding of a subject or natural phenomena, while applied research is oriented toward practical knowledge with economic value, agbiotech research can be viewed along a continuum from the very basic to that directly applied to the farm. Furthermore, both public and private sectors conduct both types of research. Nevertheless, the private sector engages primarily in research that creates economically significant knowledge, while most scientific research focusing on relatively abstract concepts is conducted in the public sector (Jaffe, 1999). Moreover, as discussed above, biotech firms pursue basic research mainly to develop their "absorptive capacity" for exploiting basic scientific knowledge generated by the public sector, and the public sector's "applied" research tends to be located closer to the basic end of the basic-to-applied continuum than is the private sector's applied research. Therefore, I will assume that private biotech firms invest primarily in applied research, while the public sector (including universities, government agencies, and federally financed R&D centers, but mainly universities) performs primarily basic research.

Practical measures of the knowledge generated in basic and applied research, and of a paper trail characterizing the information flows between these two research programs, will be discussed in detail in the empirical model chapter. With the help of these practical measures, we will be able to examine how new scientific knowledge is produced from universities' bioscience research, how this scientific knowledge spills into biotech firms and interacts with the research inputs at those firms to produce applied agbiotech innovations, and how firms' applied research efforts spill back into universities' basic research. A visual description of these knowledge flows is given in figure 3.1.

3.2 Knowledge Production Function Framework

As discussed in the previous chapter, dual models of optimal agricultural R&D investment in generally imperfectly competitive market structures have been

Figure 3.1 Information Flows Between Universities' Basic Bioscience Research and Biotech Firms' Applied Biotechnology Research



developed by, among others, Moschini and Lapan (1997), Oehmke et al. (1999), and Koo and Wright (1999). Consistent with Foltz, Barham, and Kim (2000) and Graff, Rausser, and Small (2001), the present study focuses on a primal model of knowledge production, first sketched and utilized by Pakes and Griliches (1980), to examine the patent-R&D expenditure relationship in non-agricultural sectors. In particular, my model stresses the relationship between basic bioscience and applied biotechnology research.

In a given time interval, let

Y_a^f be the outputs of applied agricultural biotech research (agbiotech innovations) at a given biotech firm;

Y_b^u the outputs of basic bioscience research (new bioscience knowledge) utilized by the given biotech firm's innovations at a given university;

K_a^f, L_a^f the physical capital quantity and biotechnologist time, respectively, employed in applied biotechnology research at the firm;

K_b^u, L_b^u the physical capital quantity and bioscientist time, respectively, employed in basic bioscience research at the university;

X^f the vector of fixed factors of the firm producing the biotechnological innovations; and

X^u the vector of fixed factors of the university producing the new bioscience knowledge.

Thinking of the research process as a classic production process using labor, capital, and other inputs, the technology of biotechnological change can be represented by

$$Y_a^f = F_a^f\ (K_a^f, L_a^f, Y_b^u, X^f) \tag{3.1}$$

$$Y_b^u = F_b^u\ (K_b^u, L_b^u, K_a^f, L_a^f, X^u) \tag{3.2}$$

where time subscripts and lag operators are suppressed for notational simplicity.

Equation (3.1) says that the number of agbiotech innovations in a given time interval depends on the quantity of capital and technologist time allocated to applied research at the biotech firm, on the level of basic bioscience knowledge which the biotech firm has the capacity to absorb, and on the biotech firm's fixed factors and characteristics that affect its research infrastructure. Variable Y_b^u in this equation reflects the fact that the informational outputs of basic research are "intermediate inputs," which are indispensable in the downstream applied research eventually leading to agbiotech innovations.

Equation (3.2) says that the quantity of new bioscience knowledge is a function of the quantity of capital and bioscientist time devoted to basic bioscience research at the unviersity, of the quantity of inputs allocated at private firms to applied research in that field, and of the university's fixed factors and characteristics. Applied research inputs K_a^f and L_a^f in this equation represent the feedback from applied research to basic bioscience. Successes of applied agbiotech research — marketable outputs such as genetic material, lab equipment and processes, proprietary gene-sequence data bases, and computational software for gene data analysis — are extensively utilized by university bioscientists to reduce the costs or increase the effectiveness of their basic research. However, failures as well as successes in applied research help stimulate fundamental questions and bring interesting phenomena to the fore, guiding bioscientists in their search for new knowledge about molecular and genetic biology. For this reason, I utilize applied research efforts K_a^f and L_a^f rather than applied research outputs Y_a^f to represent the feedbacks, in equation (3.2), from applied to basic reserach.

Equations (3.1) and (3.2) provide a model of the influence of universities' basic bioscience research investments on commercial agbiotech innovations. The universities employ scientists and capital in their basic biological research programs. After the appropriate lags, these investments generate bioscience outputs Y_b^u at a rate depending upon the universities' fixed factors such as their location, history, and institutional and qualitative characteristics. Biotech firms employ scientists, technologists, and capital resources to exploit the bioscience outputs and to develop them into economically significant agbiotech innovations. The success rate of this effort depends also on the firms' fixed factors such as their location, size, market orientation, management strategy, and other characteristics.

Analogous to a producer's conventional optimization behavior, let us assume the biotech firm and university choose (K_a^f, L_a^f) and (K_b^u, L_b^u) to maximize their applied research output Y_a^f and basic research output Y_b^u, respectively, given their respective research budget constraints:

$$\underset{K_a^f, L_a^f}{Max} \quad F_a^f \;(K_a^f, L_a^f, Y_b^u, X^f)$$

$$s.t. \quad (W_k)_a^f \, K_a^f \; + \; (W_l)_a^f \; L_a^f \; \leq \; E_a^f \tag{3.3}$$

$$\underset{K_b^u, L_b^u}{Max} \quad F_b^u \;(K_b^u, L_b^u, K_a^f, L_a^f, X^u)$$

$$s.t. \quad (W_k)_b^u \, K_b^u \; + \; (W_l)_b^u \; L_b^u \; \leq \; E_b^u \tag{3.4}$$

where

$(W_k)_a^f$ denotes the rental price of physical capital employed in applied research at the firm producing the biotechnological innovations;

$(W_l)_a^f$ the price of biotechnologist time employed in applied biotechnological research at the firm;

$(W_k)_b^u$ the rental price of physical capital employed in basic research at the university generating the new bioscience knowledge;

$(W_l)_b^u$ the price of bioscientist time employed in basic bioscience research at the university;

E_a^f expenditures on applied biotechnological research at the firm; and

E_b^u expenditures on basic bioscience research at the university.

At the optimal allocations of (K_a^f, L_a^f) and (K_b^u, L_b^u) derived from the first-order conditions of the above two mazimizations, I obtain

$$Y_a^f = G_a^{f'} \left(\frac{(W_k)_a^f}{E_a^f}, \frac{(W_l)_a^f}{E_a^f}, Y_b^u, X^f \right) \tag{3.5}$$

$$Y_b^u = G_b^{u'} \left(\frac{(W_k)_b^u}{E_b^u}, \frac{(W_l)_b^u}{E_b^u}, K_a^f, L_a^f, X^u \right) \tag{3.6}$$

Note that (K_a^f, L_a^f) cannot be replaced by E_a^f in equation (3.6) because the university cannot control the research input allocation at the biotech firm. Let us further assume competitive markets for scientific labor and physical capital, so that $(W_k)_a^f = (W_k)_b^u$ and $(W_l)_a^f = (W_l)_b^u$. Holding these prices fixed, equations (3.5) and (3.6) can be rewritten as

$$Y_a^f = G_a^f \left(E_a^f, Y_b^u, X^f \right) \tag{3.7}$$

$$Y_b^u = G_b^u \left(E_b^u, K_a^f, L_a^f, X^u \right) \tag{3.8}$$

Since basic research outputs rarely produce direct economic benefits, but instead are necessary "immediate inputs" to applied research, a society might wish in general to allocate research inputs so as to maximize long-run technological innovation, which contributes directly to technical change, productivity growth, and economic development. Of course, such decisions in the United States are not centrally controlled. However, understanding the structure of our agbiotech research system and the principal forces driving agricultural biotechnological change, can under certain intellectual property protection schemes and industry structures provide useful guidance to the allocation of public research funding.

Suppose basic bioscience output Y_b^u is an intermediate input and agbiotech innovation Y_a^f is maximized as the final output, subject to a total research budget constraint $E^0 = E_a^f + E_b^u$. Substituting (3.8) into (3.7) in this maximization gives us

$$\underset{E_a^f, E_b^u, K_a^f, L_a^f}{Max} \quad G_a^f \left[E_a^f, \; G_b^u \left(E_b^u, K_a^f, L_a^f, X^u \right), \; X^f \right]$$

$$s.t. \quad (W_k)_a^f \, K_a^f + (W_l)_a^f \, L_a^f \leq E_a^f$$

$$E_a^f + E_b^u \leq E^0 \tag{3.9}$$

The socially optimal levels of (E_a^f, E_b^u) and (K_a^f, L_a^f) can be derived from the first-order conditions of this maximization. At the optimal levels of such choice variables, the maximized Y_a^f is obtained as

$$Y_a^f = H' \left[(W_k)_a^f , (W_l)_a^f , E^0 , X^f , X^u \right] \tag{3.10}$$

As the socially optimal level of (K_a^f, L_a^f) is determined simultaneously with the optimal allocation of (E_a^f, E_b^u), we can get rid of the associated prices and rewrite equation (3.10) as

$$Y_a^f = H \left(E^0 , X^f , X^u \right) . \tag{3.11}$$

For the same reason, basic bioscience research output equation (3.8) can be expressed in terms of expenditures on basic and applied research:

$$Y_b^u = G_b^u \left(E_b^u , E_a^f , X^u \right) \tag{3.8'}$$

Recall applied biotechnological research output equation (3.7):

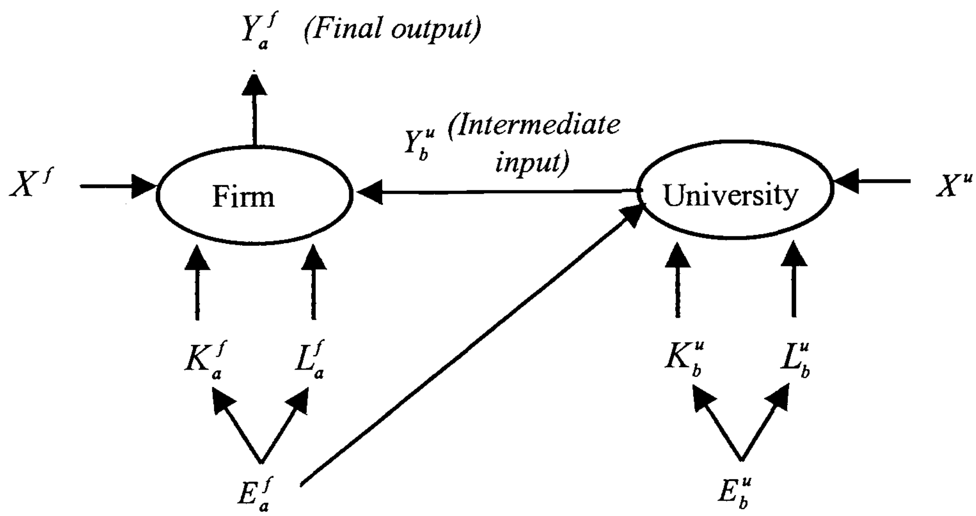
$$Y_a^f = G_a^f \left(E_a^f , Y_b^u , X^f \right) \tag{3.7}$$

The relationships between research investments and the increments in the stocks of useful scientific or technological knowledge in equations (3.7) and (3.8') are termed knowledge production functions, or sometimes research production functions. The spillovers from basic to applied, and from applied to basic, research in these equations provide the essentials for testing whether the two enterprises are complements or substitutes for one another. Because allocations of research inputs, especially of the capital input in equations (3.1) and (3.2), are difficult to observe, and because labor and capital inputs may not be separable in either basic or applied research process, equations (3.7) and (3.8') form the principal model to be estimated in the present study. Substituting (3.8') into (3.7) gives the applied research output reduced form

$$\begin{aligned} Y_a^f &= G_a^f \left(E_a^f, G_b^u \left(E_b^u, E_a^f, X^u \right), X^f \right) \\ &= G_a^f \left(E_a^f, E_b^u, X^u, X^f \right) \end{aligned} \tag{3.12}$$

This model is summarized in figure 3.2.

Figure 3.2 Summary of the Knowledge Production Network



$$(W_k)_a^f K_a^f + (W_l)_a^f L_a^f \leq E_a^f \qquad\qquad (W_k)_b^u K_b^u + (W_l)_b^u L_b^u \leq E_b^u$$

$$E_a^f + E_b^u \leq E$$

In the derivation of the estimable knowledge production functions above, I first maximized Y_a^f and Y_b^u over (K_a^f, L_a^f) and (K_b^u, L_b^u), respectively. I then optimized jointly over (K_a^f, L_a^f) and (E_a^f, E_b^u) on behalf of the social planner. The same results can be obtained by first optimizing jointly over (K_a^f, L_a^f) and (K_b^u, L_b^u) on behalf of the social planner, then socially optimizing again over (E_a^f, E_b^u). The derivation following this approach is presented in Appendix A.

3.3 Complementarity Between Basic and Applied Research

Holding X^f and X^u fixed, the total differential of (3.12) is

$$dY_a^f = \frac{\partial G_a^f}{\partial E_a^f} dE_a^f + \frac{\partial Y_a^f}{\partial Y_b^u} \frac{\partial Y_b^u}{\partial E_b^u} dE_b^u + \frac{\partial Y_a^f}{\partial Y_b^u} \frac{\partial Y_b^u}{\partial E_a^f} dE_a^f \tag{3.13}$$

Suppose a social budget restriction is enforced over the basic and applied research programs. That is, $E^0 = E_a^f + E_b^u$, implying $dE_a^f = -dE_b^u$. Any decrease in the applied research budget equals the corresponding increase in the basic research budget. Substituting the social budget restriction into (3.13) and divding both sides by dE_b^u gives the total derivative

$$\frac{dY_a^f}{dE_b^u} = -\frac{\partial Y_a^f}{\partial E_a^f} + \frac{\partial Y_a^f}{\partial Y_b^u}\left(\frac{\partial Y_b^u}{\partial E_b^u} - \frac{\partial Y_b^u}{\partial E_a^f}\right) \tag{3.14}$$

This total derivative of the applied research output reduced form gives us the total effect on applied research of reallocating another scarce research dollar to basic research from applied reserch.

The total effect has direct and indirect components. The direct effect, represented by the first right-hand term in the above equation, is that of reducing the output rate of applied biotech innovations. In addition, as reflected in $-\partial Y_b^u / \partial E_a^f$, the reallocation reduces the rate at which bioscience knowledge is generated to the extent that applied research activity, which the budget cutback has retarded, stimulates basic bioscience. Through the positive effect of basic bioscience on applied agbiotech innovations, $\partial Y_a^f / \partial Y_b^u$, this reduction feeds back by reducing the rate of applied innovations. On the other hand, the extra dollar spent on bioscience research increases basic research output by $\partial Y_b^u / \partial E_b^u$, which then increases the rate of applied agbiotech innovations by way of the positive influence $\partial Y_a^f / \partial Y_b^u$. The latter two effects on applied innovations are called indirect because they are realized through the influence of basic science on applied innovations instead of through a direct change in applied research expenditures. If $\left(\partial Y_a^f / \partial Y_b^u \right) \left(\partial Y_b^u / \partial E_b^u \right)$ exceeds $\left(\partial Y_a^f / \partial E_a^f \right) + \left(\partial Y_a^f / \partial Y_b^u \right) \left(\partial Y_b^u / \partial E_a^f \right)$ in absolute value, equation (3.14) is positive, implying that reallocating money from applied to basic research increases the applied innovation rate. Such will occur only if the impact of basic research expenditures on basic research outputs, $\partial Y_b^u / \partial E_b^u$, is especially large.

We now are able to distinguish between bioscience's partial and total effect on agricultural biotechnological innovations. The partial effect, $\partial Y_a^f / \partial Y_b^u$, presumably is positive because scientific insights usually enhance the returns to biotechnological research effort. But total effect $d\, Y_a^f / d\, Y_b^u$ is positive only if (3.14) is, since only then do Y_a^f and Y_b^u both rise as basic bioscience research budget E_b^u does. That is, basic and applied research outputs in the agbiotech field are complementary if both rise as research funding is exogenously reallocated toward basic science research.

In a diagram of a production possibility frontier with basic scientific knowledge and applied technological innovations as the two outputs, the allocation represents a movement along a frontier on which total research budget is held fixed.

Complementarity occurs wherever the frontier has positive slope. Bioscience in these zones acts more as an input to, than as a competing output with, agricultural biotechnology. Where there are agglomeration or network economies, rational research administrators likely operate within such complementarity zones. But in the presence of decreasing returns, decision makers might push beyond such zones and operate where the two research outputs are substitutes for one another.

A knowledge production function of this sort differs from conventional ones in which inputs are rival and tradable. Basic research generally has been conducted following the tradition of "open science", which offers complete, rapid, and free disclosure of results and methods. In that case, a firm buys informational outputs from basic research at zero market price but at a shadow cost equaling the resources expended to evaluate and exploit the information. Such a cost depends on the firm's fixed factors and accumulated science know-how.

The sample variation necessary for estimating model (3.7) and (3.8') is provided by inter-year and inter-firm or inter-university differences in research expenditures and productivity. Hence, the model in this study will reflect a weighted average of individual firms' and universities' knowledge-production technologies. Nevertheless, it can readily be used to draw inferences about the effects of individual firms' and universities' R&D expenditures (E_a^f and E_b^u) and characteristics (X^f and X^u), and spillovers between them. These will enable one to test hypotheses about the principal factors affecting the production of basic bioscience knowledge and agricultural biotechnological innovations, and the synergy between them.

Chapter 4: Empirical Modeling and Initial Testing

Based on the knowledge production function framework sketched above, I now specify estimation equations for biotech firms' applied biotechnology research and universities' basic bioscience research. I then discuss important issues in empirical model specification and estimation, such as the practical measures of research outputs and information flows, and selection of functional form, lag structure, and estimation technique.

4.1 Practical Measures of Research Outputs and Information Flows

4.1.1 Measures of Applied and Basic Research Outputs

The first task in specifiying an empirical model is to identify practical measures of research outputs. This is not an easy exercise because research activities generate intangible knowledge assets, whose values are not as easy to measure as are tangible outputs produced from a conventional production process. Nevertheless, consistent with most recent literature, I will use patent awards as the measure of applied research output and scientific publications as the measure of basic research output.

The strengths and weaknesses of using patents as indicators of increments to economically useful knowledge resulting from applied research have been discussed extensively in Griliches (1990) and Cockburn, Henderson, and Stern (1999). A patent is a property right granted by a government agency to exclude others from making, using, or selling the invention for a limited time (in the United States currently 20 years from the date on which the application for the patent is filed), in exchange for public disclosure of the invention when the patent is granted. Among the three types of patent presently granted in the U.S., the utility patent, design patent, and plant patent[1], utility patents provide the strongest and broadest protection to inventors. According to the U.S. Patent and Trademark Office (USPTO), "Utility patents may be granted to anyone who invents or discovers any new and useful process, machine, article of manufacture, or compositions of matters, or any new useful improvement thereof." The patent law specifies that, in order for an invention to be patentable, it must be novel, non-obvious, and useful, and that mere ideas and suggestions cannot be patented. Hence, patent awards by definition relate to a firm's or industry's inventiveness and reflect the increment to it's stock of economically useful knowledge. This is especially true for biotechnology — the urge to patent innovations has become part of the culture of this industry since patent protection was extended to living organisms in the early 1980's. Moreover, patents are quantitative and readily accessible, perhaps the most readily accessible, indicator of inventions, which make them attractive in empirical economic analysis.

However, a patent measure does have problems, the major ones being that not all inventions are patentable or patented, and that patents differ greatly in their technical and economic significance. For example, some biotech firms, especially large ones, retain certain discoveries as trade secrets, developing and marketing them on their own rather than patenting and licensing them to other entities. In this case,

[1] A design patent protects new or original design for an article of manufacture. A plant patent may be granted to an invention or discovery and asexual reproduction of any new and distinct variety of plant.

the patent measure understates the firms' research productivity. On the other hand, many biotech firms' patented discoveries perform unsuccessfully in the subsequent field-trial phase of R&D, or even if they are successful there, prove later to be commercially unprofitable. Indeed, by surveying patent owners, investigating patent renewal patterns, or modeling the relationship between patents and market value, economists have found very high variances and skewnesses in the distribution of patent values.

The problem that not all inventions are patented can, one would think, be at least partly taken care of by limiting the analysis to a particular industry, namely the biotech industry in the present study, as all firms in the same industry probably share similar research characteristics and patenting strategy and culture. But variations in patent quality have led to doubts about the appropriateness or accuracy of patent counts as an indicator of firms' applied research outputs. Fortunately, the U.S. Patent Office requires patent applicants to cite the earlier-patented inventions which the applicants' own inventions have utilized. Thus, a patent's future citations reflect how broad and technologically important the cited patent turned out to be and may also be suggestive of the royalty or licensing revenues that may be generated by the cited patent. This leads to a frequent way of accounting for patent quality: to weight each patent by the number of citations it eventually receives from subsequent patents. However, there is an inherent difficulty with the citation-weighted measure: truncation bias, namely, that patents awarded in later years have had less opportunity than those in earlier years to be cited, so that citation rates of later-awarded patents tend to be downward-biased. Truncation bias likely is especially serious in the agricultural biotech industry, given the comparatively short time window available for patent data in this infant industry. For this reason, in the present study, I have to employ raw patent counts in the econometric analysis.

When we turn to the measurement of outputs from basic academic bioresearch, the first candidate that comes to our mind is scientific publications. The public-good nature of basic science research makes it impossible to establish proprietary rights over research results or to appropriate benefits derived from them. But a non-market-based incentive for producing public good knowledge has been in place for a long time: by being the first to communicate an advance in scientific knowledge, a scientist can establish priority of discovery, a form of intellectual property right bringing recognition from the scientific community instead of direct financial benefit. Disclosing research findings in a timely fashion in scientific journals is a necessary step in establishing this priority. The key role played by publication counts in determining researchers' salaries, promotions, and tenure status reconfirms the importance of publication in academia. Furthermore, published articles are quantitative and easily accessible, adding to their appeal as direct measures of the increments to basic scientific knowledge.

One of the most frequently encountered criticisms of raw publication counts as research output measures is that they do not adequately account for quality differences. Academic researchers widely agree that published papers vary significantly in scientific quality. A paper authored by a prominent scientist and published in a top field journal might be widely cited, while a paper in a lower-ranked journal might not attract much attention. Since cited publications are believed to be a useful input in a field's future publication output, scientific publications can be quality-weighted in a fashion similar to patents, namely by accounting for the number of times they are cited in subsequent publications. Unfortunately, data on subsequent publication citation performance of the scientific papers included in our dataset are not available.[1] Instead, we do have access to the number of times a given paper was cited by subsequent agbiotech patents. Although this number understates the overall scientific impact of the given paper, it is specifically indicative of the paper's influence in agricultural biotechnology. Moreover, preliminary evidence suggests that bioscience publications highly cited in publicly accessible scientific literature tend also to be highly cited in biotech patents (CHI Research). Hence, in the context of the present study, the number of citations to a bioscience publication made from subsequent agbiotech patents is a robust proxy of the quality of that publication. Publication counts weighted by this quality measure also are subject to citation truncation bias, for the same reason a citation-weighted patent measure is. Therefore, raw counts of scientific publications are used in the empirical estimation in the present study.

Let $AgPt_{j\tau}$ represent the number of agricultural biotechnology patents awarded to the j^{th} firm in the t^{th} year, and $Science_{it}$ be the number of bioscience

[1] Although it is feasible to collect citations to each published paper made from subsequent papers by manually searching each subsequent year's Science Citation Index and summing the annual citation counts for the paper in question, the time required by this procedure prevented us from doing so.

publications authored by bioscientists at the i^{th} university in the t^{th} year. These are unweighted measures of firms' applied and universities' basic research outputs. If we define $A_{j\tau}^{\tau'}$ as the number of times in the τ'^{th} year ($\tau' = \tau+1, \tau+2, \ldots, T_\tau$) that a patent document cites any of the agbiotech patents awarded to the j^{th} firm in the τ^{th} year, $B_{it}^{\tau^*}$ as the number of times in the τ^{*th} year ($\tau^* = t+1, t+2, \ldots, T_t$) that a patent document cites any of the bioscience publications authored by bioscientists at the i^{th} university in the t^{th} year, and r_s as the discount rate in the s^{th} year, the quality-weighted measures of research outputs will be

$$AgPtCite_{j\tau} = \sum_{\tau'} A_{j\tau}^{\tau'} \;/\; \left(1 + r_{\tau'-\tau}\right) \tag{4.1}$$

$$ScienceCite_{it} = \sum_{\tau^*} B_{it}^{\tau^*} \;/\; \left(1 + r_{\tau^*-t}\right) \tag{4.2}$$

These are, respectively, the time-discounted number of patent citations to firm j's τ^{th}-year agbiotech patents and to university i's t^{th}-year published bioscience articles, summed over the years following the award of the cited patents or the publication of the cited papers. The literature on the economics of technology and science has recognized the potential noise in patent and publication data as measures of research outputs, although they are the best measures we currently have.

4.1.2 Measures of Information Flow Between Basic and Applied Research

Having identified the measures of applied and basic research output, I now construct a practical measure of information flow between basic and applied research.

Recognizing that the output of basic research rarely possesses intrinsic economic value, but instead is a critically important input to the downstream research investment that yields technological innovations, economists have tried to link universities' basic research to private firms' applied research in an explicit manner. But efforts to trace the use of basic research have been frustrated by the lack of any generally reliable means to quantify the use of the informational outputs of a given basic research program. For example, as reviewed in sections 2.3.3.1 and 2.4.1.1, Jaffe (1989) modeled university research as an input to private firms' patent production and allowed for endogenous determination of university and industry R&D expenditures at the state level. Audretsch and Stephan (1996) examined the institutional connection of university-based scientists with biotech firms and the importance of various roles these scientists played in affecting the geographical proximity between scientist and firm. Zucker, Darby, and Brewer (1998) studied the contribution of "star" scientists, who discovered important gene sequences, to the growth and location of the biotech industry. However, these frameworks did not require detailed origin and destination data for information flows, so none were able to identify the mechanism through which knowledge produced in basic university research was transported to firms' technological innovations.

Although David, Mowery, and Steinmueller (1992) and Griliches (1994) have noted that citations from patents to the scientific literature could be useful for tracing information flows from basic to applied research (sometimes called the "knowledge externalities of basic research"), no empirical measures of this sort have been reported. In the present study, I develop an empirical and explicit paper trail of information flows between basic bioscience and applied agricultural biotechnology. The trail is developed by recording each citation that each agbiotech patent made to each bioscience paper. Let $C_{j\tau}^{it'}$ be the number of times that agbiotech patents awarded to the j^{th} firm in the τ^{th} year cite any of the bioscience papers authored by scientists at the i^{th} university in the t'^{th} year ($t' = \tau_0, \ldots, \tau-2, \tau-1$). The universities' scientific informational outputs that feed into a biotech firm's agbiotech patent production can be expressed as

$$SciInput_{j\tau} = \sum_{t'} \sum_{i} C_{j\tau}^{it'} \; / \; \left(1 + r_{\tau - t'}\right). \tag{4.3}$$

In equation (4.3), Σ_i is the number of times that all the universities' t'^{th} year bioscience publications were cited by firm j in year τ. Summing this over all t' years gives the time-discounted total quantity of scientific inputs into the applied agbiotech research program of the given firm in the given year.

Constructed this way, measure (4.3) allows one to trace applied research outcomes back to the antecedent basic research. It consequently facilitates the tracing of patent-producing firms back to their associated paper-producing universities. To my knowledge, this is the first successful effort to do so.

4.2 Other Outputs, Research Inputs, and Characteristics Variables

In the knowledge production function framework, my main interest is in the relationship between R&D expenditures and research outputs, and in the interactions between basic bioscience and applied biotechnological research. Other characteristics and fixed factors, however, might affect research productivity.

4.2.1 Other Outputs, Research Inputs, and University Characteristics

A university's major functions are to conduct research and provide education. An essential part of a bioscience graduate education is the experimental techniques and research expertise gained from working in university research labs. Hence, the number of graduate students in agricultural and biological sciences, denoted here by *Grad*,[1] can serve as a measure of a university's education output, which is jointly produced with its research output, namely scientific publications. On the other hand, graduate students are also a university input because they not only receive education from but provide scientific labor to the university's research programs. After graduation, the intellectual human capital built from their graduate education will move from the home university to some other university or to a private firm, providing

[1] All the variables discussed in this section, except those representing university fixed factors, are for university *i* in year *t*, where the university and time subscripts are suppressed for notational simplicity. No time subscript is needed for fixed factors, as they remain constant over time.

a source of knowledge or information spillover to other institutions. This spillover effect is difficult to trace and is not modeled in the present study.

Universities' agbiotech-related science research falls mainly in the life science field, which consists of the agricultural, biological, and medical science disciplines. Although medical science accounts for the biggest share of life science R&D expenditure at most universities, and some techniques developed in medical research can be used for agricultural purposes, the majority of research in medicine does not have agricultural applications. For this reason, R&D expenditures and other inputs in medical science contain more noise than useful information in the evaluation of agbiotech-related basic research. Hence, I focus on the agricultural and biological fields only.

Agricultural and biological science involves substantial access to lab equipment and research materials. The overwhelming importance of physical resources to the research process makes research and development (R&D) expenditure the most significant measure of a university's capacity to publish scientific articles and to provide research training to graduate students. Possible measures of R&D expenditure in the agricultural and biological disciplines include: (a) *AgRDtot*, *BioRDtot*, *AgRDfed*, and *BioRDfed*, total and federally-sponsored R&D expenditures in agricultural and biological sciences, respectively, and (b) *USDARD*, *NIHRD*, and *NSFRD*, R&D expenditures provided by the U.S. Department of Agriculture (USDA), National Institutes of Health (NIH), and National Science Foundation (NSF), respectively.[1] The above two sets of measures may overstate agbiotech-related basic research budgets to the extent that the indicated research constitutes only a portion of that financed by the measured R&D expenditure. On the other hand, they may understate the budgets to the extent that some agbiotech-related basic research expenditures are reported in other disciplines or are provided by other agencies. Unfortunately, a precise match between research input and output data can be obtained only from a survey at the individual research project level. Yet the present measures are the most disaggregated and best proxies that are readily available. Since NIH and NSF sponsor research in many different fields, R&D measures referring to the "agricultural and biological sciences" appear to be the most accurate indicator of the university-agbiotech-related science research budget.

Postdoctoral fellows have become necessary inputs to life science research. It is a common practice, for example, for a molecular biologist to have two or more post-docs working in his lab. The numbers of post-docs in agricultural and biological sciences are denoted here by *AgPD* and *BioPD*, respectively. To proxy the quality of university scientists in these two disciplines, I have developed measures of agriculture- and biology-related graduate program quality, denoted by *AgRank* and *BioRank*, respectively. As discussed in more detail in section 5.2.1, a quality rank number is given only to the top 40 to 50 programs in each discipline. All other programs are assigned a rank number of 114, which is the median between 51 and 177, the total number of universities included in our sample.

[1] Conversations with university biologists suggests that the major agricultural and biological research funding sources are USDA, NIH, and NSF.

To proxy the feedback from private-sector applied research to university basic research, I include aggregate agricultural R&D expenditures (*PrivAgRD*) in the private sector as an explanatory variable. This variable can be interpreted as measuring the spillins from private agricultural research to public research.

A university's characteristics include its overall size, its faculty quality, the composition of its research funds, the university's orientation toward life science research, the strength of applied research in the university's life science research, and the research and teaching intensities of its faculty. A university's overall size can be measured by *Faculty* — the total number of faculty, or *Enroll* — the total enrollments. Average faculty salary, *FacSalary*, can be used to capture overall faculty quality. A university's research funds generally derive from five sources: the federal government, state government, industry, the institution itself, and other sources. The composition of its research funds can be represented by the share of funds from each source (*FedRD*, *StateRD*, *IndRD*, *InstRD*, and *OtherRD*). The ratio of a university's R&D expenditures in the life sciences to its total R&D expenditures, denoted by *LifeRDshare*, reflects its orientation toward life science research. The number of agbiotech patents awarded to a university, denoted by *UnivAgPt* can represent its strength in applied research. Two measures, average R&D expenditures and average student enrollment per faculty, are employed here to measure a university's research and teaching intensity, respectively. That is, *ResearchIntense=TotalRD/Faculty* and *TeachIntense=Enroll/Faculty*. All such characteristics describe the general structure and environment in which a university's agricultural and biological science research is conducted.

A university's basic science research is also affected by relatively fixed factors that change little over time. These include:

LandGrant	a dummy variable equaling one for a land-grant university and zero otherwise;
BEARegion	eight dummies to representing the eight BEA economic regions, one if a university is located in the indicated region and zero otherwise;
MedSchool	a dummy variable equaling one if the university has a medical school and zero otherwise;
PrivPub	a dummy variable equaling one for a private university and zero for a public university;
HighDegree	four dummies to represent the four types of highest degree awarded (doctorate, master's, bachelor's, and two-year associate degrees), one if that degree is the highest the university is authorized to award and zero otherwise.

I can now specify the econometrically estimable versions of (3.8'), using the unweighted measure of basic bioscience research output:

$$Science = G^{u}\ (\ Grad,\ AgRD,\ BioRD,\ AgPD,\ BioPD,\ AgRank,\ BioRank,\ PrivAgRD,\ Charact^{u},\ X^{u},\ t\ ) \tag{4.4}$$

where $Charact^u$ and X^u represent the charateristics variables and fixed factors, respectively, described above. Time trend t is included to control for changes over time in the knowledge production function or in scientists' propensity to publish. Time subscripts and lag operators will be provided in the following section when the lag structure is discussed.

Assuming input-jointness in the production of scientific publications and graduate students, we have also

$$Grad = G^{u'}\,(Paper, AgRD, BioRD, AgPD, BioPD, AgRank, BioRank$$

$$PrivAgRD, Charact^u, X^u, t\,). \tag{4.5}$$

Including both outputs in these specifications will allow estimation of the university's production possibility frontier.

4.2.2 *Other Outputs, Research Inputs, and Biotech Firm Characteristics*

Biotechnology is a new industry that is knowledge-based and composed predominantly of small start-up firms and large, established ones. Most start-up firms are formed to develop the promising inventions of academic scientists. Once such efforts are successful, they typically are acquired by established firms and the production and marketing of most final products are undertaken by the latter. Hence, most start-up companies focus on R&D activities in a limited number of research lines. In contrast, established firms typically have multiple product lines and devote only part of their R&D effort to biotechnology. Our use of a tight filter in identifying agricultural biotech patents, as discussed in section 5.1.1, exclude from the variable

Patent many biotech patents that have no immediate agricultural applications. Thus, in addition to *AgPatent*, the variable *NonAgPatent* was constructed to represent the annual numbers of patents issued to the relevant firm in all fields other than agricultural biotechnology. This includes patents in all non-biotech fields and those in biotech fields but without agricultural applications.

As shown by Hausman, Hall, and Griliches (1984), Griliches (1990), and others, R&D expenditure is the most significant available indicator of a firm's capacity to develop patented technologies. Inasmuch as patent-R&D relationships vary across technology areas, one ideally would need a firm's R&D expenditure by research field or technology class in order to analyze the contribution of R&D to its patent production. Hence, R&D expenditure data in the present study should, to the extent possible, be specific to agbiotech field. In the industry R&D database, R&D expenditures related to agbiotech include those in chemistry, biology, inorganic chemicals, agricultural chemicals, and food and tobacco. Unfortunately, most firms do not in most years report R&D expenditure by field. Instead, they report overall R&D expenditure, denoted here by *FirmRD*. Although such a measure may not precisely correspond to the firms' agbiotech patent outputs, it is the most disaggregated R&D measure readily available. Moreover, controlling for the firms' non-agbiotech patent outputs permits us to measure the impact of the overall R&D budget on these two categories of patent production.

Scientists and engineers are the labor input in a firm's patent production process. In the knowledge-based biotech industry, scientists' and engineers' intellectual capital play a specially important role in innovation: scientists need not only conduct applied research, but have the capacity to understand and explore advances in basic bioscience. The total number of scientists and engineers is here denoted by SE.

Because of the science-based nature of the biotech industry, universities' scientific information outputs serve as important inputs to biotech firms' patent production. An empirical measure of information flow from university to firm was constructed in section 4.1.2, providing an explicit paper trail between the two by tracing citations from patents to scientific papers. That measure, $SciInput$, is included as an explanatory variable in firms' agbiotech patent production functions.

A firm's characteristics, such as its overall size, its R&D intensity, and the composition of its R&D expenditures, indicate the structure and environment in which the firm conducts its research. A firm's size can be measured by its total sales ($Sales$) or total number of domestic employees ($Employee$). The ratio of a firm's total R&D expenditures to its total sales, denoted by $RD\,int$, and the ratio of a firm's scientist and engineer numbers to its total employee numbers, denoted by $SE\,int$, can be used to capture its R&D intensity. The composition of the firm's R&D can be represented by the shares of its funding derived from alternative sources such as its own or federal funds, or by the shares of its R&D allocated to alternative functions, such as basic research, applied research, and development. However, like R&D expenditure data by research field, those by funding source and function are reported only by few firms in few years. Finally, firms' geographical locations are denoted by *BEARegion*.

Employing the unweighted measure of applied agbiotech research output, the econometrically estimable version of (3.7) can be specified as

$$AgPt \;=\; G^{f}\ \left(NonAgPt,\, SciInput,\, FirmRD,\, SE,\, Charact^{f},\, X^{f},\, t \right) \tag{4.6}$$

where $Charact^{f}$ and X^{f} refer to the charateristics variables and fixed factors described above. Time trend t is included to control for changes over time in biotech firms' knowledge production functions or their propensity to patent. The time subscripts and lag operators are suppressed for notational simplicity.

Assuming input-jointness in the two applied research outputs, agbiotech patents and non-agbiotech patents, we have also

$$NonAgPt \;=\; G^{f'}\ \left(AgPt,\, FirmRD,\, SE,\, Charact^{f},\, X^{f},\, t \right) \tag{4.7}$$

As data are not available on the allocation of scientific inputs to non-agbiotech patent production, variable $SciInput$ is not included in equation (4.7). Any bias in this omission likely is not serious because private-sector R&D in most fields is not as science-based as it is in biotechnology. Similar to the university model, a production possibility frontier between the two outputs can be estimated using this specification.

4.3 Functional Form

Now that we have discussed the output and input variables useful in the knowledge production functions, we require a functional form to characterize the specific relationship between research outputs, R&D expenditures, and other inputs.

In principle, only one functional form may reflect the true data-generating process. But in practice, the data-generating process is unknown, and decisions about functional form usually are made according to research objectives and data availability.

In their pioneer empirical study of firm-level R&D-to-patent relationships, Pakes and Griliches (1980) found the form best fitting their data was a modified Cobb-Douglas, namely the logarithm of patents as a function of the logarithm of current and past R&D expenditures. In many later studies (Jaffe 1989; Pardey 1989; and Adams and Griliches 1996), the double-log form continued to be appealing, although others, such as simple linear and log-linear forms, have also been used (Foltz, Kim, and Barham 2001 and Graff, Rausser, and Small 2001). All these inflexible forms have the advantage of being easy to estimate and parsimonious in parameters, and of permitting simple computations of economic effects (Alston, Norton, and Pardey 1995). But such advantages come at a cost: the inflexible functional forms are restrictive. For example, the Cobb-Douglas function imposes constant output elasticity and constant unitary substitution elasticities among inputs.

Flexible functional forms have been developed that do not impose a degree of economy or scale and that allow a full range of substitution elasticities. The most commonly used are the translog and Generalized Leontief. The disadvantages of these forms are that they require many more parameters than do their inflexible counterparts, in turn consuming substantial degrees of freedom and exacerbating multicollinearity problems. Moreover, complicated functional forms often nurture implausible implications in the results, many of which are difficult to detect.

Although flexible functional forms have gained popularity in the last two decades, inflexible forms have been used in most econometric studies of knowledge production, since they require relatively few parameter estimates, conserve degrees of freedom, and help to avoid multicollinearty. In this, the rather weak data availability in knowledge production relationships has been the overriding consideration.

I conducted a preliminary examination of functional forms. Regressions using a quadratic form, relating research outputs to linear and quadratic forms of all inputs and to interaction terms for some of them, revealed significant multicollinearity. Most parameter estimates were not significant. Consistent with the knowledge production literature, this suggested that an inflexible functional form be used. Since a significant number of zeros are present in both output and input data at both the university and firm, logarithms cannot be taken. Therefore, a simple linear form for the dependent and input variables is employed in this study. Curvature in the production possibility relationships between a university's or firm's outputs is captured by the use of both linear and square root terms for the right-hand-side output variables.

4.4 Lag Structure

An important issue in the empirical estimation of knowledge production functions is the structure of the research gestation lag, that is, the lag between the inception and completion of a research project. In practice, this is the lag between research inputs and outputs. While the lag structure of an individual research project might be identifiable, aggregate analysis of input-output relationships in many research projects, institutions, and years requires assumptions about the shape and length of the R&D gestation lag.

Alston, Norton, and Pardey (1995) summarize a wealth of evidence on three types of lag structure in agricultural research. The first involves well-defined, deterministic, and finite lag structures, such as the inverted-V, trapezoidal, and polynomial. They have been used to reduce the number of estimated coefficients, but it is questionable whether a finite lag process can completely capture the effects of past research and whether any single prior about lag structure shape is appropriate. The second type of lag structure allows for greater flexibility by pretesting the lag length and weights within a specified lag form, or by searching the lag weights provided by probability-generating functions such as the binomial and Pascal distributions, then choosing those minimizing the sum of squared residuals. Although such probability distributions are quiet flexible[1], the final choice of distribution form imposes a definite shape on the lag structure. A third approach has been to use a form-free lag, by including current and lagged research expenditures as separate explanatory variables in the knowledge production function. This approach is appealing if one is interested in the sum of the lagged effects and in the mean lag distribution, but it still requires a presumption about the lag length and cannot guarantee non-negativity of coefficients on current and lagged R&D expenditures.

---

[1] Different parameterization of binomial and Pascal distributions can approximate various distributions, such as the normal, geometric, Poisson, and Gaussian, and different skewnesses of those distributions.

Recent empirical applications of the knowledge production mostly have employed the form-free lag structure, using short to medium lag lengths from one year upwards (Acs and Audretsch, 1988; Crepon and Duguet, 1997) to seven (Kim, 1999). Other studies (Jaffe, 1989; Graff, Rausser, and Small, 2001) have employed only the current inputs or lagged research expenditures in one previous year. After extensive empirical work, little consensus has yet been achieved on the appropriate form of research lag.

4.4.1 Lag Structure in University Model

Following the form-free lag approach, I tested various lag lengths (one, two, and three years) in the university bioscience production function. None provided good fits and lag coefficients were estimated with low precision. I then made an attempt to regress a two-year moving average of bioscience paper output against a one-year lagged, two-year moving average of all the independent variables. The intention was to smooth out the randomness in the correspondence between R&D spending and science output. However, this approach did not produce better estimates.

The weak explanatory power of the above specifications suggests that some earlier R&D expenditures, which continue to affect a university's bioscience production, were being omitted in the model. Pardey and Craig (1989) find that the lag in public agricultural research output can be as long as 30 years. Considering that our university science model characterizes basic research in the agricultural and biological sciences, and that basic research is believed on average to take longer than applied research does, a relatively long lag length is needed. Yet the goodness-of-fit of longer lag lengths cannot be tested with the short data horizon (1985-1997) we have available.

In conclusion, I employ a compact parametric model in the university science model by including a lagged dependent variable on the right-hand side, essentially incorporating infinite lags for all regressors but requiring a small number of parameters. Following this approach, equation (4.5) can be rewritten as

$$\begin{aligned} Science_{it} = {} & \alpha_0 + \lambda\, Science_{i,t-1} + \alpha_1\, Grad_{it} + \alpha_2\, Grad_{it}^{0.5} + \\ & \alpha_3\, AgRD_{it} + \alpha_4\, BioRD_{it} + \alpha_5\, AgPD_{it} + \alpha_6\, BioPD_{it} + \\ & \alpha_7\, AgRank_{it} + \alpha_8\, BioRank_{it} + \alpha_9\, PrivAgRD_{it} + \\ & \alpha_{10}\, Charact_{it}^{u} + \alpha_{11}\, X_i^{u} + \alpha_{12}\, t \qquad (4.8) \end{aligned}$$

where subscripts i and t index universities and years, respectively. The absolute value of the coefficient on the lagged dependent variable should not exceed one; otherwise, the associated dynamic process is unstable. While allowing for infinite lags, this geometric distributed lag model assigns arbitrarily small weights to the distant past; specifically, the effect of a regressor in a given year on current science production is proportionate to the effect of that regressor in the previous year by a factor of λ.

In the equation relating university graduate education to university inputs, the dependent variable I employ is the number of graduate students currently enrolled instead of the number of graduate degrees currently awarded. Since currently enrolled graduate students receive their education from the resources spent in the current year, no lag should exist between inputs and outputs in this equation. Thus, it can be expressed in the following functional form with i, t subscripts defined as before:

$$\begin{aligned} Grad_{it} \;=\; & \beta_0 + \beta_1\, Science_{it} + \beta_2\, Science_{it}^{0.5} + \\ & \beta_3\, AgRD_{it} + \chi_4\, BioRD_{it} + \beta_5\, AgPD_{it} + \beta_6\, BioPD_{it} + \\ & \beta_7\, AgRank_{it} + \beta_8\, BioRank_{it} + \beta_9\, PrivAgRD_{it} + \\ & \beta_{10}\, Charact_{it}^{u} + \beta_{11}\, X_i^{u} + \beta_{12}\, t \end{aligned} \tag{4.9}$$

For simplicity, equations (4.8) and (4.9) will be called the science and graduate student equations, respectively.

4.4.2 Lag Structure in Biotech Firm Model

The biotech firm model consists of two knowledge production functions in which a firm's agbiotech and non-agbiotech patents respectively depend on its total R&D expenditures, other research inputs, and characteristics, and in which the two research outputs are affected by one another. As discussed in Chapter 5, our patent data end in the 2000 award year, whereas biotech firms' R&D data run only through 1998. Thus, a minimum of a two-year lag had to be imposed on biotech firms' research inputs in order to include the patents awarded in 1999 and 2000. This is especially important given that the number of agbiotech patents awarded increased exponentially in the late 1990's.

Tests of various lag structure specifications similar to those in the university science model were conducted for biotech firms' agbiotech patent production. Results were quite robust to alternative lag lengths employed, such as a two-year, three-year, or four-year lag. But regressing a two-year moving average of patent quantities against a two-year lagged, two-year moving average of all the independent variables reduced the statistical significance of the coefficient on R&D expenditures. Including a lagged dependent variable on the right-hand side of these equations did not significantly improve the goodness-of-fit, implying a finite lag between the inception and completion of agbiotech research projects would be more appropriate.

The latter is consistent with the fact that private firms perform mainly applied research, which generally requires shorter lag lengths than does basic research. Moreover, a large number of start-up companies in the agricultural biotechnology industry remain for only a few years before being merged into or acquired by established companies. Indeed, most larger companies did not begin agbiotech research until recently. These facts argue for a comparatively short lag in the biotech firm agbiotech patent production model. Because the typical lag between application for and issuance of an agbiotech patent is 2.5 to 3.0 years and the patent application date is relatively close to the time when the research is performed, a three-year lag is employed on all biotech firm input variables. Hence, equation (4.6) can be re-expressed as

$$\begin{aligned} AgPt_{j\tau} \; = \; & \gamma_0 + \gamma_1 \, NonAgPt_{j\tau} + \gamma_2 \, NonAgPt_{j\tau}^{0.5} + \\ & \gamma_3 \, SciInput_{j\tau} + \gamma_4 \, FirmRD_{j,\tau-3} + \gamma_5 \, SE_{j,\tau-3} + \\ & \gamma_6 \, Charact_{j,\tau-3}^{f} + \gamma_7 \, X_j^f + \gamma_8 \, t \end{aligned} \tag{4.10}$$

where j and τ refer to the j^{th} firm in the τ^{th} year.

Symmetric to the specification of the firm's agbiotech patent production function, the non-agbiotech patent production equation (4.7) can be rewritten as

$$\begin{aligned} NonAgPt_{j\tau} \;=\; & \varphi_0 \;+\; \varphi_1\, AgPt_{j\tau} \;+\; \varphi_2\, AgPt_{j\tau}^{0.5} \;+ \\ & \varphi_4\, FirmRD_{j,\tau-3} \;+\; \varphi_5\, SE_{j,\tau-3} \;+ \\ & \varphi_6\, Charact_{j,\tau-3}^{f} \;+\; \varphi_7\, X_j^f \;+\; \varphi_8\, t \end{aligned} \tag{4.11}$$

For simplicity, equations (4.10) and (4.11) will be called agbiotech and non-agbiotech equations hereafter, respectively.

4.5 Estimation Issues

In order to implement the university science model consisting of equations (4.8) and (4.9), and the biotech firm's patent model consisting of equations (4.10) and (4.11), several estimation issues need to be addressed.

4.5.1 Count Data Model

When analyzing patent-R&D relationships, a commonly used dependent variable is a count of the total number of patents applied for or issued to a particular firm in a given year. Count data approaches have been developed to explicitly reflect the non-negative integer nature of this dependent variable and the associated preponderance of zeros and small values. In these approaches, Poisson or negative binomial distributions are used to represent the average counts of events (patents) that occur both "randomly and independently" over a period of time, with distribution parameters related to the firm's input and characteristics variables. Refinements and adaptations have been made to handle over-dispersion, weak exogeneity, individual effects, and serial correlation (Hausman, Hall, and Griliches, 1984; Trivedi, 1997; Crepon and Duguet, 1997).

As discussed in Chapter 5, a "fractional count" approach has instead been used in the present study, whereby a fraction of one credit is assigned to a paper author institution or patent awardee firm according to its share in the list of authors or awardee institutions. This counting method should be more appropriate than count-data approaches in measuring research output quantities, especially as inter-institutional collaborations are common in agricultural biotechnology. Hence, the dependent variables in our model -- the quantities of bioscience papers and quantities of agbiotech and non-agbiotech patents awarded -- are constructed as continuous variables and count-data assumptions are not applicable to our case. The implications of converting our "fractional counts" to the nearest integers, then estimating a count data model, deserves further research but will not be covered in this dissertation.

4.5.2 Fixed vs. Random Effects Model

The most contentious issue in the use of panel data, such as the pooled time-series, cross-university or cross-firm data employed in the present study, concerns the nature of the time-specific and section-specific variables. While time-specific effects are proxied by simple linear time trend variables t, university or firm research productivities may differ in ways the included regressors cannot completely explain.

Two frameworks can be employed to capture these university- or firm-specific effects: the fixed effects and random effects models.

The fixed effects approach assumes the unobservable differences across sections (universities or firms) can be captured in a section-specific constant term, which does not change over time. For example, the fixed effects model for biotech firms' agbiotech patent production can be formulated as

$$\begin{aligned} AgPt_{j\tau} \;=\; & \gamma_j + \gamma_1 \, NonAgPt_{j\tau} + \gamma_2 \, NonAgPt_{j\tau}^{0.5} + \\ & \gamma_3 \, SciInput_{j,\tau} + \gamma_4 \, FirmRD_{j,\tau-3} + \gamma_5 \, SE_{j,\tau-3} + \\ & \gamma_6 \, Charact_{j,\tau-3}^{f} + \gamma_7 \, X_j^f + \gamma_8 \, t + \varepsilon_{j\tau} \end{aligned} \tag{4.12}$$

in which γ_j is the individual firm's effect to be estimated and $\varepsilon_{j\tau}$ is the random disturbance term reflecting the inherently stochastic nature of the research process. Model (4.12) can be estimated by applying OLS, including dummy variables for each section. A drawback is that the bulk of the variance in panel data usually is in the cross-sectional dimension, so that a fixed effects model neglects much of the information in the data and often leads to nonsignificant estimates. Moreover, if the assumption is not true that the section-specific effect is constant over time, the fixed-effect estimates are biased.

The random effects model instead treats the unobserved section-specific effect as a random variable by incorporating, in addition to the conventional residual, a disturbance term characterizing each individual section. This disturbance term has constant probability moments. The random effects model of biotech firms' agbiotech patent production can be formulated as

$$
\begin{aligned}
AgPt_{j\tau} \;=\; & \gamma_0 + \gamma_1\, NonAgPt_{j\tau} + \gamma_2\, NonAgPt_{j\tau}^{0.5} + \\
& \gamma_3\, SciInput_{j\tau} + \gamma_4\, FirmRD_{j,\tau-3} + \gamma_5\, SE_{j,\tau-3} + \\
& \gamma_6\, Charact_{j,\tau-3}^{f} + \gamma_7\, X_j^f + \gamma_8\, t + \mu_j + \varepsilon_{j\tau} \qquad (4.13)
\end{aligned}
$$

where μ_j is the disturbance for the j^{th} firm with mean zero, variance σ_u^2, and no correlation with any $\varepsilon_{j\tau}$. Consistent and efficient Generalized Least Squares (GLS) estimators have been developed to estimate this model. Some have argued that the individual effect should always be treated as random. However, no justification exists for treating the individual effect as uncorrelated with the other explanatory variables, as is assumed in the random effects model. So parameter estimates in the random effects model may be inconsistent.

Since there is little theoretical guidance about whether unobservable section-specific effects are present, and if they are, whether the fixed effects or random effects approach is appropriate, these questions are taken up in our empirical estimations below.

4.5.3 Contemporaneous Correlation and Endogeneity

As the university and firm model each consists of two equations characterizing the production of two outputs, one would expect correlations to exist between the disturbance terms in the two sets of equations. If correlations are present, OLS estimation of each equation is still consistent but not efficient. Alternatively, estimating the two equations in a single model, as an equation system using Seemingly Unrelated Regressions (SUR), would provide us with efficient estimates (Greene 1997). The greater the correlation between the disturbances is, the greater the efficiency gain accruing to SUR. Comparisons between OLS and SUR estimates are provided in Chapter 6.

Because each of the two equations in each model contains one output as the dependent variable and the other output as an explanatory variable, the two output variables are endogenous in appearance, requiring simultaneous equation estimation such as Two Stage Least Squares (2SLS) or Three Stage Least Squares (3SLS). However, a closer examination of model structure shows that each of the two equations represents maximization of a separate output, so that each corresponds to a separate maximand or institutional culture. Hence, as shown below, the two outputs are not endogenously determined and the two equations are not simultaneous.

4.6 Production Possibility Frontiers

A production possibility frontier (PPF) can be derived from each of the two equations, (4.8) and (4.9), in the university model and from each of the two equations, (4.10) and (4.11), in the biotech firm model. In the present section, I discuss the construction of a single PPF from each equation pair.

Consider first the university model. The paper equation shows the maximum number of cited bioscience publications that the representative university can achieve given a fixed number of graduate students, while the graduate student equation shows the maximum number of graduate students the university can achieve given a fixed number of cited bioscience publications. In the biotech firm model, the agbiotech equation reflects the maximum number of agricultural biotechnology patents given the representative biotech firm can achieve a fixed number of non-agbiotech patents awarded. The non-agbiotech equation reflects the maximum number of non-agbiotech patents the firm can achieve given fixed levels of agbiotech patents. Thus, in each model, the two equations represent two optimization processes, each representing a separate portion of a production possibility frontier.

This can be illustrated for a biotechnology firm using the generalized PPF curve shown in figure 4.1. The PPF shape in figure 4.1 can be characterized by an output directional distance function:

$$\vec{D}\left(X, Y, g_Y \right) = Max \left\{ \beta : \left(Y + \beta g_Y \right) \in P\left(X \right) \right\} \tag{4.14}$$

where Y is the output set consisting of a biotech firm's two outputs (agbiotech and non-agbiotech patents), X is the firm's input set, g_Y is the output directional vector, and P represents the production technology. When g_Y is defined as (1, 0) and (0, 1) in agbiotech and nonagbiotech patent outputs, the directional distance function can be rewritten respectively in these two outputs as

$$\vec{D}_{AgPt}\left(X, AgPt, NonAgPt; \left(1, 0 \right) \right) =$$

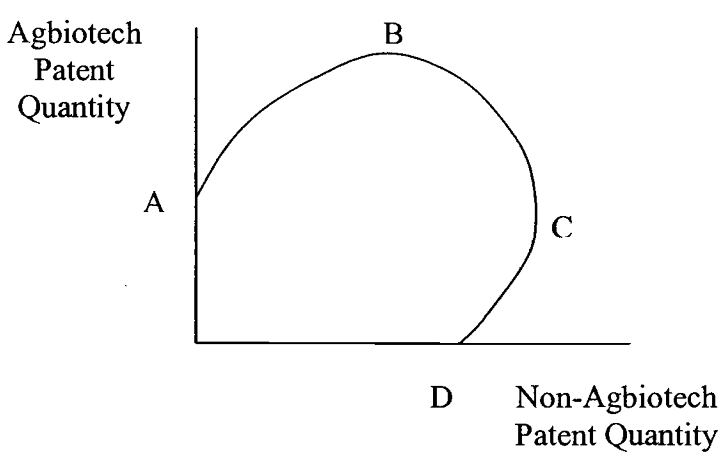
$$Max \left\{ \beta_{AgPt} : \left(AgPt + \beta_{AgPt} ,\ NonAgPt \right) \in P\left(X \right) \right\} \tag{4.15}$$

$$\vec{D}_{NonAgPt}\left(X\,,AgPt\,,NonAgPt\,;(0\,,1\,)\right) =$$

$$Max\ \left\{\beta_{NonAgPt}\ :\ \left(AgPt\ ,\ NonAgPt+\beta_{NonAgPt}\right)\in\ P\left(X\right)\right\} \qquad (4.16)$$

The above two distance functions are alternative representations of the agbiotech equation (4.10) and non-agbiotech equation (4.11). The PPF shape in figure 4.1 cannot be characterized in a single estimable regression context, because two agbiotech patent quantities correspond to a range of fixed non-agbiotech

Figure 4.1 Generalized Production Possibility Frontier: Illustration for a Biotechnology Firm



quantities, and two non-agbiotech quantities correspond to a range of fixed agbiotech quantities.

To see this more clearly, observe how the PPF in figure 4.1 is traced out. Starting from point A, the agbiotech-maximizing firm reallocates a fixed quantity of input from the production of agbiotech to non-agbiotech patents. At first, both agbiotech and non-agbiotech outputs may rise because the technical insights gained in producing non-agbiotech innovations may act as an input to agbiotech innovations. As reallocation proceeds, the two products become substitutes at point B: additional units of non-agbiotech are now less an input to than a substitute for agbiotech. As the fixed input quantities continue to be reallocated to point C, the agbiotech maximizer encounters the problem that both outputs start to decline. It would be inoptimal for the firm to operate in zone CD. Suppose instead that a firm maximizing non-agbiotech innovations begins from point D. Reallocating the fixed inputs at first increases both agbiotech and non-agbiotech patents, so the firm would continue moving in this direction. At point C, the two outputs become substitutes. Depending on the relative prices of the two outputs, it still is potentially rational to operate in this zone. However, at point B, both outputs start to decline and it would be inoptimal for the non-agbiotech maximizer to operate in zone BA.

At least two types of firms therefore are represented in figure 4.1: the agbiotech maximizers, at the top of the PPF (ABC), are the firms that specialize in agricultural biotechnology and which therefore can be assumed to maximize the number of agricultural biotechnology innovations permitted by a given set of fixed resources. The non-agbiotech maximizers, at the right side of the PPF (BCD), are those firms that specialize mostly in pharmaceutical or other technologies and which therefore would maximize those outputs at given input levels. For these firms, agbiotech is a sidelight. We therefore need two equations to characterize the PPF in figure 4.1. The equations are not simultaneous, as no endogeneity exists between the two outputs, but instead reflect different corporate cultures or R&D strategies. A similar analysis applies to the university model, where universities can alternatively have a research emphasis or a teaching emphasis.

Returning again to the biotech firm model, it would be desirable if both the agbiotech and non-agbiotech maximizers' technologies could be represented in a single PPF, i.e., in a single "grand-technology." However, it is impossible to do so in a regression context for the very reason that the PPF as shown in figure 4.1 cannot be characterized by a function. The alternative is to estimate two separate equations and, for each, to trace out a PPF for a given level of inputs, namely by predicting values of the left-hand-side output variable at various levels of the right-hand-side one. The common set of these two PPFs constitutes an approximation of the grand production possibility frontier. The grand-technology is fairly accurately estimated in this fashion between A and B (for agbiotech maximizers), and between C and D (for non-agbiotech maximizers). But it likely is poorly represented between B and C, where a splicing of the two equations must occur.

Some functional forms might be better than others at approximating given portions of this grand PPF. For example, a quadratic function is better able to accommodate the negatively sloped plus one positively sloped portion of figure 4.1, but it would not necessarily perform well in either of the positively sloped PPF portions taken alone. The virtue of the square-root functional form, employed in this dissertation, is that its curvature is greatest at lower levels of the outputs. This makes it better at approximating the positively sloped portions of the PPF where, because of the right-skewed nature of our sample (see Chapter 6), the bulk of the firm and university data lie.

4.7 Optimal Allocation of R&D Expenditures, Returns to R&D, and Other Elasticities

4.7.1 Optimal Allocation of R&D Expenditures

In the theoretical model of the knowledge production function derived in Chapter 3, I assume an optimal allocation between inputs such as capital and labor, and include total R&D expenditures as an explanatory variable instead of breaking them down by input category. In the empirical university model, however, the number of post-docs in ag and bio sciences are included as explanatory variables along with R&D expenditures. The cost of post-docs is, of course, a part of R&D expenditures. Let *AgO* and *BioO* denote aggregate non-post-doc R&D inputs in agricultural and biological sciences, respectively, at the representative university, such as faculty research time, equipment, and genetic materials used in experiments. Then total university agricultural and biological R&D expenditures can be expressed as

$$AgRD = W_{AgPD}\ AgPD + W_{AgO}\ AgO \tag{4.17}$$

$$BioRD = W_{BioPD}\ BioPD + W_{BioO}\ BioO \tag{4.18}$$

where W_{AgPD} and W_{BioPD} are post-doc salaries and W_{AgO} and W_{BioO} are the aggregate prices of non-post-doc R&D inputs in the ag and bio sciences, respectively.

Similarly, the number of scientists and engineers is included in the empirical biotechnology firm model along with the firms' total R&D expenditures. Let *NonSE* denote firms' aggregate R&D inputs other than scientists and engineers, consisting primarily of research equipment, buildings, and supplies. Then total R&D expenditures at the firm can be expressed as

$$FirmRD = W_{SE}\ SE + W_{NonSE}\ NonSE \qquad (4.19)$$

where W_{SE} and W_{NonSE} are scientist and engineer salaries and price of non-salary inputs, respectively.

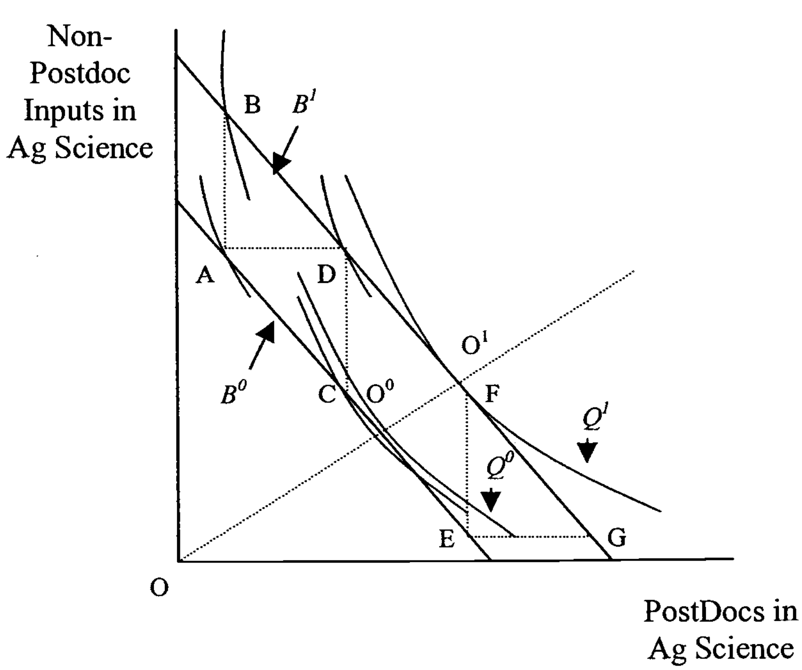
This formulation of the empirical models allows us to test the optimal allocation of R&D expenditure between post-docs and all other R&D inputs at universities, and that between scientists and engineers and all other R&D inputs at biotech firms. The production of bioscience with two agricultural science inputs, *AgPD* and *AgO*, is characterized in figure 4.2 for illustration. If total agricultural R&D expenditure *AgRD* is optimally allocated between those two inputs, the university would operate along expansion path $O\text{-}O^0\text{-}O^1$. For example, under the *AgRD* budget constraint represented by isocost curve B^0, the maximum number of bioscience papers the university can produce is represented by isoquant Q^0, which is tangent to isocost curve B^0 at point O^0. At this point, the allocation between the two R&D inputs is cost-minimizing, so output would not change if the university moves away from point O^0 along isocost curve B^0 by an indefinitely small amount. However, if the university allocates the two inputs inoptimally, such as at point A to the left of the optimal allocation, where too much is spent on non-post-doc inputs, a movement along the isocost curve to the southeast would increase the university's bioscience output. In contrast, if the university operates to the right of the optimal allocation, such as at point E, a movement along the isocost curve toward the northwest would increase it's bioscience output.

This effect on bioscience production of a movement along an isocost curve can be captured by the coefficient of *AgPD* in equation (4.8), that is by

$$\alpha_5 = \frac{\partial\ Science}{\partial\ AgPD}\bigg|_{AgRD^0} \tag{4.20}$$

the partial derivative of bioscience output with respect to *AgPD*, holding total agricultural R&D expenditure and all other variables in equation (4.8) fixed. Thus,

Figure 4.2 University Production of Bioscience with Two Agricultural Science Inputs



when $\alpha_5 = 0$, the university's agricultural R&D is optimally allocated between post-docs and non-post-doc inputs. If $\alpha_5 > 0$ or $\alpha_5 < 0$, the university operates to the left or right of the optimal allocation, and too few or too many post-docs are employed, respectively.

Similarly, parameter β_5 in equation (4.8) can be used to test the optimality of the allocation between post-docs and non-post-doc inputs in the university's biology program:

$$\beta_5 = \left. \frac{\partial \, Science}{\partial \, BioPD} \right|_{BioRD^0} \tag{4.21}$$

Following the same logic, parameters γ_5 and φ_5 in equations (4.10) and (4.11), respectively representing agbiotech and non-agbiotech patent production, indicate whether the representative biotechnology firm's R&D expenditures are optimally allocated between scientists/engineers and non-salary inputs:

$$\gamma_5 = \left. \frac{\partial \, AgPt}{\partial \, SE} \right|_{FirmRD^0} \tag{4.22}$$

$$\varphi_5 = \left. \frac{\partial \, NonAgPt}{\partial \, SE} \right|_{FirmRD^0} \tag{4.23}$$

4.7.2 Returns to R&D

A major interest of our study is the effect of R&D expenditure on research output as measured by scientific publications or patents, in another words, the returns to R&D. As inputs here are specified in linear form, parameter estimates indicate marginal effects. Multiplying marginal effects by the corresponding ratios of input to output quantities gives the elasticities of research output with respect to research input evaluated at selected data points. In the following section, measures of marginal effects and elasticities of R&D expenditures are developed in order to analyze R&D impacts on university and firm knowledge production under alternative policy settings.

4.7.2.1 Returns to University R&D

Returns to scale in university agricultural science R&D are discussed here for illustration. Their counterpart in the biological sciences can be derived analogously.

Suppose under a fixed agricultural R&D budget B^0, the university originally operates at inoptimal point A, as shown in figure 4.2. Now permit the budget to increase by one unit to B^1. The marginal effect on bioscience output of this increase in *AgRD* can be decomposed into the vertical movement from A to B, or the horizontal movement from A to D, or any combination of these two movements between B and D, depending upon the university's flexibility in distributing this extra unit of agricultural R&D expenditure between post-docs (*AgPD*) and non-post-doc inputs (*AgO*).

The one-unit increase in *AgRD* consists of expenditure increases in the two R&D inputs,

$$\begin{aligned} \partial\, AgRD &= \partial \left(W_{AgPD}\, AgPD + W_{AgO}\, AgO \right) \\ &= W_{AgPD}\, \partial AgPD + W_{AgO}\, \partial AgO \end{aligned} \tag{4.24}$$

where time and university subscripts are suppressed for simplicity. If the university can spend this extra unit of research money only on non-post-doc inputs, the number of post-docs will remain at its original level. The marginal effect of this increase in the university's bioscience production is then

$$\frac{\partial\, Science}{\partial\, AgRD}\Bigg|_{AgPD^0} = \frac{\partial\, Science}{W_{AgO}\ \partial\, AgO}\Bigg|_{AgPD^0}$$

$$= \frac{1}{W_{AgO}}\ \frac{\partial\, Science}{\partial\, AgO}\Bigg|_{AgPD^0} = \alpha_3 \qquad (4.25)$$

This marginal effect is measured by the coefficient of *AgRD*, i.e., parameter α_3, in the university's science equation (4.8) and corresponds to the vertical movement from A to B in figure 4.2.

As shown in equation (4.17), the slope of *AgPD*, parameter α_5, in equation (4.8) gives the marginal effect on bioscience output of a one-unit increase in agricultural science post-docs, holding total agricultural R&D expenditure fixed. This marginal effect can be translated into expenditure terms by incorporating the price of post-docs, as follows:

$$\frac{\partial\, Science}{W_{AgPD}\ \partial\, AgPD}\Bigg|_{AgRD^0} = \frac{1}{W_{AgPD}}\ \frac{\partial\, Science}{\partial\, AgPD}\Bigg|_{AgRD^0} = \frac{1}{W_{AgPD}}\ \alpha_5 \qquad (4.26)$$

But since $AgRD = W_{AgPD}\, AgPD + W_{AgO}\, AgO$ is fixed, we have

$W_{AgPD}\, \partial\, AgPD = -\, W_{AgO}\, \partial\, AgO$, that is, the cost of non-post-doc inputs must fall by as much as the cost of post-docs rises. Therefore, (4.26) can be expressed as

$$\frac{1}{W_{AgPD}} \frac{\partial\, Science}{\partial\, AgPD} \bigg|_{AgRD^0} =$$

$$\frac{\partial\, Science}{W_{AgPD}\, \partial\, AgPD} \bigg|_{AgO^0} - \frac{\partial\, Science}{W_{AgO}\, \partial\, AgO} \bigg|_{AgPD^0} \qquad (4.27)$$

Such a marginal movement along the budget line is composed of two effects. The first right-hand term represents the effect of increasing post-doc expenditure by one unit, holding non-post-doc inputs fixed. The second term gives the effect of reducing non-post-doc expenditure by one unit, holding the number of post-docs fixed. These two effects correspond respectively to the horizontal movement from A to D and vertical movement from D to C in figure 4.2.

The horizontal movement from A to D represents a situation in which the university spends its entire marginal dollar on post-docs, keeping its non-post-doc inputs at their original level. The effect of this one-unit increase in agricultural R&D expenditure on the university's bioscience production can in this situation be derived by rearranging (4.27) and substituting (4.25) and (4.26) into it as follows,

$$\frac{\partial\, Science}{\partial\, AgRD} \bigg|_{AgO^0} = \frac{1}{W_{AgPD}} \frac{\partial\, Science}{\partial\, AgPD} \bigg|_{AgO^0}$$

$$= \frac{\partial\ Science}{W_{AgPD}\ \partial\ AgPD} \Big|_{AgRD^0} + \frac{\partial\ Science}{W_{AgO}\ \partial\ AgO} \Big|_{AgPD^0}$$

$$= \frac{1}{W_{AgPD}} \alpha_5 + \alpha_3 \tag{4.28}$$

Equations (4.25) and (4.28) offer two polar approaches by which the university may invest an extra unit of agricultural R&D expenditure, namely, by a vertical or a horizontal movement in figure 4.2. If the university chooses to expend a combination of the two inputs, the total marginal effect would be an average of the two polar cases, weighted by the corresponding expenditure share of each input in this extra unit of R&D expenditure. This average effect corresponds to a movement from A to any point between B and D in figure 4.2. The point at which the university arrives depends upon the combination it chooses. That is, the average marginal effect of one more unit of agricultural R&D is computed as

$$\frac{\partial\ Science}{\partial\ AgRD} \Big|_{Average} =$$

$$\frac{\partial\ Science}{\partial\ AgRD} \Big|_{AgPD^0} \left(1 - \frac{W_{AgPD}\ \partial\ AgPD}{\partial\ AgRD} \right) + \frac{\partial\ Science}{\partial\ AgRD} \Big|_{AgO^0} \left(\frac{W_{AgPD}\ \partial\ AgPD}{\partial\ AgRD} \right)$$

$$= \alpha_3 \left(1 - \frac{W_{AgPD}\ \partial\ AgPD}{\partial\ AgRD} \right) + \left(\frac{1}{W_{AgPD}} \alpha_5 + \alpha_3 \right) \left(\frac{W_{AgPD}\ \partial\ AgPD}{\partial\ AgRD} \right)$$

$$= \alpha_3 + \alpha_5 \frac{\partial\ AgPD}{\partial\ AgRD} \tag{4.29}$$

Note that, while the marginal effect of the horizontal movement depends on choice of post-doc salary, the average marginal effect in (4.29) does not.

As R&D expenditures and post-doc employment in biological science are included in the university model symmetrically to those in agricultural science, the vertical, horizontal, and average marginal effects of a one-unit increase in biological R&D can be derived following the same approach used for agricultural R&D. Specifically,

$$\frac{\partial\, Science}{\partial\, BioRD}\bigg|_{BioPD^0} = \alpha_4 \tag{4.30}$$

$$\frac{\partial\, Science}{\partial\, BioRD}\bigg|_{BioO^0} = \frac{1}{W_{BioPD}}\alpha_6 + \alpha_4 \tag{4.31}$$

$$\frac{\partial\, Science}{\partial\, BioRD}\bigg|_{Average} = \alpha_4 + \alpha_6 \frac{\partial\, BioPD}{\partial\, BioRD} \tag{4.32}$$

If a university receives one more unit of R&D funding and invests it proportionately to the observed average R&D expenditures in these two disciplines, the marginal effect of the increase would be the average of the marginal effects in the agricultural and biological sciences weighted by their respective shares in this extra unit of total R&D spending:

$$\frac{\partial\, Science}{\partial\, AG\,\&\,BioRD}\bigg|_{Average} = \frac{\partial\, Science}{\partial\, AgRD}\bigg|_{Average}\left(\frac{AgRD}{AgRD + BioRD}\right)$$

$$+ \frac{\partial\, Science}{\partial\, BioRD}\Bigg|_{Average}\left(\frac{BioRD}{AgRD + BioRD}\right) \tag{4.33}$$

All the marginal effects developed here are short-run measures. As a lagged dependent variable is included in university science equation (4.8) to allow for infinite lags, the long-run marginal effects of the R&D rise equal the corresponding short-run measures multiplied by $1/(1-\lambda)$.

Having developed above the marginal effects of R&D expenditures under alternative assumptions about how the university uses the extra unit of R&D, we may transform the marginal effects into elasticities by multiplying them by the corresponding ratios of R&D expenditure to output quantity. For example, when a university invests an extra unit of biological science R&D in both post-docs and non-post-doc inputs, the short- and long-run elasticities of university bioscience output with respect to R&D expenditure in a given year are, respectively:

$$E^{S.R.}_{Science-BioRD}\Big|_{Average} = \frac{\partial\, Science}{\partial\, BioRD}\Bigg|_{Average}\left(\frac{BioRD}{Science}\right)$$

$$= \left(\alpha_4 + \alpha_6 \frac{\partial\, BioPD}{\partial\, BioRD}\right)\left(\frac{BioRD}{Science}\right) \tag{4.34}$$

$$E^{L.R.}_{Science-BioRD}\Big|_{Average} = \left(\frac{\partial\, Science}{\partial\, BioRD}\right)^{L.R.}\Bigg|_{Average}\left(\frac{BioRD}{Science}\right)$$

$$= \left(\frac{1}{1-\lambda}\right)\left(\alpha_4 + \alpha_6 \frac{\partial\, BioPD}{\partial\, BioRD}\right)\left(\frac{BioRD}{Science}\right) \tag{4.35}$$

Diminishing, constant, or increasing returns predominate when the above elasticities are greater than, equal to, or smaller than unity.

4.7.2.2 Returns to Biotechnology Firm R&D

As a firm's scientist and engineer numbers are included along with its total R&D expendture in agbiotech and non-agbiotech equations (4.10) and (4.11) in the same fashion that numbers of post-docs and R&D expenditures are included in the university paper equation, the vertical, horizontal, and average marginal effects of a one-unit increase in the firm's total R&D on its agbiotech and non-agbiotech patent production can be derived following the same approach used in the university model.

Using parameter estimates from the biotech firm's agbiotech patent equation (4.10), the three marginal effect measures can be computed respectively as

$$\left.\frac{\partial\, AgPt}{\partial\, FirmRD}\right|_{SE^0} = \gamma_4 \tag{4.36}$$

$$\left.\frac{\partial\, AgPt}{\partial\, FirmRD}\right|_{NonSE^0} = \frac{1}{W_{SE}}\gamma_5 + \gamma_4 \tag{4.37}$$

$$\left.\frac{\partial\, AgPt}{\partial\, FirmRD}\right|_{Average} = \gamma_4 + \gamma_5 \frac{\partial\, SE}{\partial\, FirmRD} \tag{4.38}$$

Equations (4.36) – (4.38) correspond to cases in which the firm invests its extra unit of R&D expenditure only on equipment, only on scientist and engineers, or on a combination of the two, respectively. Similarly, marginal effects on its non-agbiotech patent output of an increase in a firm's R&D are based on parameter estimates from equation (4.11) as follows

$$\frac{\partial\ NonAgPt}{\partial\ FirmRD}\Bigg|_{SE^0} = \varphi_4 \tag{4.39}$$

$$\frac{\partial\ NonAgPt}{\partial\ FirmRD}\Bigg|_{NonSE^0} = \frac{1}{W_{SE}}\varphi_5 + \varphi_4 \tag{4.40}$$

$$\frac{\partial\ NonAgPt}{\partial\ FirmRD}\Bigg|_{Average} = \varphi_4 + \varphi_5 \frac{\partial\ SE}{\partial\ FirmRD} \tag{4.41}$$

Let *AllPt* denote a firm's total annual patent output. Then $AllPt = AgPt + NonAgPt$ and $\partial\ AllPt = \partial\ AgPt + \partial\ NonAgPt$. Thus, the increase in the firm's total patent production brought about by a one-unit increase in its R&D expenditure is the sum of the increases in its agbiotech and non-abiotech patent production brought about by that R&D increase. Specifically,

$$\frac{\partial\ AllPt}{\partial\ FirmRD}\Bigg|_{SE^0} = \frac{\partial\ AgPt}{\partial\ FirmRD}\Bigg|_{SE^0} + \frac{\partial\ NonAgPt}{\partial\ FirmRD}\Bigg|_{SE^0} \tag{4.42}$$

$$\frac{\partial\ AllPt}{\partial\ FirmRD}\Bigg|_{Equip^0} = \frac{\partial\ AgPt}{\partial\ FirmRD}\Bigg|_{NonSE^0} + \frac{\partial\ NonAgPt}{\partial\ FirmRD}\Bigg|_{NonSE^0} \tag{4.43}$$

$$\frac{\partial\ AllPt}{\partial\ FirmRD}\Bigg|_{Average} = \frac{\partial\ AgPt}{\partial\ FirmRD}\Bigg|_{Average} + \frac{\partial\ NonAgPt}{\partial\ FirmRD}\Bigg|_{Average} \tag{4.44}$$

Again, multiplying these marginal effects by the corresponding ratios of firms' R&D expenditures to research outputs gives the R&D returns in elasticity form.

4.7.3 Other Elasticities

Other elasticities also are of interest and can provide insight into the factors affecting university and firm research productivity. Those elasticities can be calculated using formulae similar to the ones described in the previous section.

In the university model, the effect of a university's agricultural and biological graduate program ranking on it's bioscience output can be derived in elasticity form by differentiating both sides of science equation (4.8) with respect to it's ranking variables and multiplying this derivative by the reciprocal of output:

$$E_{Science-AgRank} = \frac{\partial\, Science}{\partial\, AgRank}\frac{1}{Science} = \alpha_7 \frac{1}{Science} \tag{4.45}$$

$$E_{Science-BioRank} = \frac{\partial\, Science}{\partial\, BioRank}\frac{1}{Science} = \alpha_8 \frac{1}{Science} \tag{4.46}$$

Note that the program ranking variables are positive integers with non-unique base values and it is meaningless to think of these as changing by a given percentage.

The elasticities of a university's bioscience output with respect to aggregate agricultural R&D expenditures in the private sector, and to the number of agbiotech patents awarded to the university in a given year, are the respective marginal effects multiplied by the ratio of the corresponding right-hand side variable to the output quantity:

$$E_{Science-PrivAgRD} = \frac{\partial\, Science}{\partial\, PrivAgRD}\frac{PrivAgRD}{Science} = \alpha_9 \frac{PrivAgRD}{Science} \tag{4.47}$$

$$E_{Science-UnivAgPt} = \frac{\partial\, Science}{\partial\, UnivAgPt}\frac{UnivAgPt}{Science} = \alpha_{10}\frac{UnivAgPt}{Science} \tag{4.48}$$

The above two elasticities show the responsiveness of a university's bioscience output to, respectively, the applied research expenditures in the private sector and research expenditures at the university itself. Elasticities for the university's graduate education output can be derived in the same fashion, using the corresponding parameter estimates from graduate student equation (4.9).

In the biotechnology firm model, the marginal effect and elasticity of science citation variable *SciInput* in agbiotech equation (4.10) is especially informative, because it not only measures the direct contribution of scientific findings to agricultural biotechnology inventions, but also explicitly links university basic bioscience research with biotechnology firms' applied research. The increase in a firm's agbiotech patent output brought about by one more cited scientific paper is:

$$\frac{\partial\, Agpt}{\partial\, SciInput} = \gamma_3 \tag{4.49}$$

This marginal effect can be expressed in elasticity form as

$$E_{AgPt-SciInput} = \frac{\partial\, AgPt}{\partial\, SciInput}\frac{SciInput}{AgPt} = \gamma_3\frac{SciInput}{AgPt} \tag{4.50}$$

The change in a firm's propensity to patent agricultural biotechnology and non-agbiotech innovations can respectively be derived by taking the natural logarithm of equations (4.10) and (4.11) and differentiating with respect to time:

$$E_{AgPt-t} = \frac{\partial \ln AgPt}{\partial t} = \frac{\partial AgPt}{\partial t} \frac{1}{AgPt} = \gamma_8 \frac{1}{AgPt} \tag{4.51}$$

$$E_{NonAgPt-t} = \frac{\partial \ln NonAgPt}{\partial t}$$

$$= \frac{\partial NonAgPt}{\partial t} \frac{1}{NonAgPt} = \varphi_8 \frac{1}{NonAgPt} \tag{4.52}$$

The change in the firm's propensity to patent both types of innovations together is the sum of the above two elasticities, weighted by the share of each innovation type in total patent numbers:

$$E_{AllPt-t} = \frac{\partial (AgPt + NonAgPt)}{\partial t} \left(\frac{1}{AgPt + NonAgPt} \right)$$

$$= (\gamma_8 + \varphi_8) \left(\frac{1}{AgPt + NonAgPt} \right)$$

$$= E_{AgPt-t} \left(\frac{AgPt}{AgPt + NonAgPt} \right) + E_{NonAgPt-t} \left(\frac{NonAgPt}{AgPt + NonAgPt} \right) \tag{4.53}$$

4.8 Complementarity Between Basic Bioscience and Applied Biotechnology

Recall that equation (3.14) gives the total effect on applied research output of reallocating another scarce dollar from applied to basic research. Substituting for the partial derivatives in those equations the corresponding marginal effects estimated from university paper equation (4.8) and biotech firm agbiotech and non-agbiotech equations (4.10) and (4.11) provides estimates of the complementarity between applied agricultural biotechnology and basic bioscience research. As derived in the previous section, derivatives $\partial Y_a^f / \partial Y_b^u$ and $\partial Y_b^u / \partial E_a^f$ in equation (3.14) are unique in our model, namely, $\partial AgPt / \partial Science = \gamma_3$ and $\partial Science / \partial PrivAgRDe = \alpha_9$, respectively. However, the other two derivatives in equation (3.14) are captured differentially, that is depending on the policy settings.

For example, alternative assumptions can be made about how a policy planner would extract research resources from a commercial firm and reallocate them to the university sector: (a1) all the reallocated money could be provided to university agricultural research; (a2) all of it could be provided to university biological research; or (a3) it could be divided between the two. Under these alternative assumptions, $\partial Y_b^u / \partial E_b^u$ in equation (3.14) would be represented by marginal effects $\frac{\partial\ Science}{\partial\ AgRD}$, $\frac{\partial\ Science}{\partial\ BioRD}$, and $\frac{\partial\ Science}{\partial\ Ag\ \&\ BioRD}$. Further assumptions can be made about how the university allocates the extra research money between post-docs and non-post-doc inputs in agricultural or biological science. That is, (b1) all the money may be allocated to non-post-doc inputs; (b2) all of it can be allocated to post-docs; or (b3) it may be allocated to a combination of these two. In agricultural science research, namely under assumption (a1), for example, marginal effects of the above three allocation plans would be represented respectively by $\frac{\partial\ Science}{\partial\ AgRD}\Big|_{AgPD^0}$, $\frac{\partial\ Science}{\partial\ AgRD}\Big|_{AgO^0}$, and $\frac{\partial\ Science}{\partial\ AgRD}\Big|_{Average}$. Similarly, if the extra research money is provided entirely to the university's biological science research, the marginal effects of allocation schemes (b1), (b2), and (b3) would be represented respectively by $\frac{\partial\, Science}{\partial\, BioRD}\Big|_{BioPD^0}$, $\frac{\partial\, Science}{\partial\, BioRD}\Big|_{BioO^0}$, and $\frac{\partial\, Science}{\partial\, BioRD}\Big|_{Average}$.

Any research money a central planner would tax away from a biotechnology firm affects both its agricultural biotechnology and non-agbiotech patent production. Yet in our empirical modeling, funds reallocated to the university sector contribute to the production only of agbiotech-patent-cited scientific papers. To measure total complementarity between basic and applied research, derivative $\partial Y_a^f / \partial E_a^f$ in (3.14) can be represented either by $\frac{\partial\, AgPt}{\partial\, FirmRD}$ or $\frac{\partial\, AllPt}{\partial\, FirmRD}$, depending upon whether one's research interest focuses only on agricultural biotechnology or expands to non-agbiotech areas as well. Symmetric to the university model, assumptions may be made about the identity of the research input from which the firm's research money is extracted. It might be extracted: (c1) entirely from non-salary inputs, (c2) entirely from scientist and engineer employment, or (c3) from a combination of these two. If one focuses on agricultural biotechnology patents only, the marginal effects in these three cases would respectively be $\frac{\partial\, AgPt}{\partial\, FirmRD}\Big|_{SE^0}$, $\frac{\partial\, AgPt}{\partial\, FirmRD}\Big|_{Equip^0}$, and $\frac{\partial\, AgPt}{\partial\, FirmRD}\Big|_{Average}$.

Complementarity between basic bioscience and applied biotechnology research, as specified in equation (3.14), can be computed under alternative policy settings, represented by the corresponding measures of $\partial Y_b^u \,/\, \partial E_b^u$ and $\partial Y_a^f \,/\, \partial E_a^f$. For example, suppose one unit of research expenditure is extracted from a combination of the firm's spending on scientists and engineers and on non-salary inputs such as equipments, buildings, and supplies. The money is then reallocated to a combination of post-docs and non-post-doc inputs in the university's agricultural research. If I concentrate on the firm's agricultural biotechnology patent production only, the complementarity measure is:

$$\left(\frac{d\ AgPt \,\big|_{Average}}{d\ AgRD \,\big|_{Average}} \right)_{FirmAgPt-UnivAgPt} = -\ \frac{\partial\ AgPt}{\partial\ FirmRD} \,\Big|_{Average} + \frac{\partial\ AgPt}{\partial\ SciInput} \left(\frac{\partial\ Science}{\partial\ AgRD} \,\Big|_{Average} - \frac{\partial\ Science}{\partial\ PrivAgRD} \right) \tag{4.54}$$

Such a complementarity is computed by substituting equations (4.38), (4.49), (4.29), and (4.47) into (4.54):

$$\left(\frac{d\ AgPt \,\big|_{Average}}{d\ AgRD \,\big|_{Average}} \right)_{FirmAgPt-UnivAgPt} = -\left(\gamma_4 + \gamma_5 \frac{\partial\ SE}{\partial\ FirmRD} \right) + \gamma_3 \left[\left(\alpha_3 + \alpha_5 \frac{\partial\ AgPD}{\partial\ AgRD} \right) - \alpha_9 \right] \tag{4.55}$$

Using alternative measures of $\partial Y_b^u \,/\, \partial E_b^u$ and $\partial Y_a^f \,/\, \partial E_a^f$ will give us alternative values of the complementarity.

Of course, units of observation in the university model are different from those in the firm model, so that *FirmRD* and *PrivAgRD*, our two measures of applied research expenditure, are never identical to one another. For the same reason, neither are *Science* and *SciInput*, our two measures of basic research output. Total complementarity measures computed as in (4.54) remain meaningful, inasmuch as *PrivAgRD* and *SciInput* reflect a representative firm and university.

Chapter 5: Data Sources and Development

The data for the econometric model divide into research outputs (patents and scientific papers), inputs to the research process, and inputs to and production outputs at biotechnology firms. To derive observations on the research output data, I drew a large sample of agricultural biotechnology patents from the United States Patent Office database, noted the identity of the awarded firms or institutions, and observed the scientific publications cited on the patent documents. I next identified the authors of those publications and the universities or labs at which they worked at time of publication.

Because of the tedium of determining ag-biotech patents, discriminating between science and non-science references on a patent, and matching alternative forms of a scientist's or a firm's names, CHI Research of Haddon Heights, New Jersey was hired to conduct the above data search and cleaning process for this study. I then collected, from other sources such as NSF and the Bureau of Census, data on the research input allocations at the identified firms and universities, including R&D expenditures and technologists' and scientists' FTEs, and matched them with the research outputs at the same institutions, taking account of the requisite lag structure. Finally, data on production inputs and outputs at the identified biotechnology firms were drawn from the Bureau of Census and matched with the research outputs at the same firms.

5.1 Research Outputs

In this study, patents, and published scientific articles cited by those patents, are used as our respective measures of applied biotechnology and basic bioscience research outputs.

5.1.1 Patents

5.1.1.1 Search Strategy and Filter

In order to obtain an appropriate measure of agricultural biotechnology patents, we need a consistent definition of agricultural biotechnology. The USDA defines ag-biotech as "a collection of scientific techniques, including genetic engineering, that are used to create, improve, or modify plants, animals, and microorganisms for human benefit." This definition seems to include both conventional breeding and modern genetic engineering techniques. In this study I use "biotechnology" in the narrower sense of referring only to molecular genetic techniques, and define agricultural biotechnology as such techniques used to alter agricultural and food products. The definition excludes products or processes with no real direct connection to agriculture, such as plants and animals genetically modified primarily for treating human or pet diseases or for research purposes, or employed primarily as pets.

With this definition in mind, we want to collect only biotechnology patents specifically related to agriculture. Although the U.S. Patent Office's database can be searched by year, name of awarded firm, patent classification, keyword, and other criteria, the search system is inadequate by itself for determining which patents relate to agricultural biotechnology. Some innovations employing genetic engineering are not listed under the 'biotechnology' classification and cannot be identified with that keyword. Other innovations relating to agriculture are not listed as "agricultural." Still other innovations with "agriculture" as keyword are for pharmaceutical instead of agricultural use.

In collaboration with CHI Research, I developed a filter for drawing agricultural biotechnology utility patents from the U.S. Patent Office Database. To identify biotechnology patents, the filter uses certain 'biotech' classifications as well as broad keywords referring to gene-related technologies. It then chooses from them the innovations designated broadly as "agricultural," and specifically those under given agricultural product keywords. Finally, those for human pharmaceutical use, research purposes, or pets are excluded by manual culling. The filter specification takes the following form:

Step 1. Find patents under explicit biotechnology International Patent Classification (IPC) classifications, namely C12N 15/02 to 15/90.

Step 2. Find patents with transgenic or recombinant DNA keywords in titles, abstracts or claims:

transgenic, gene* near4 (transf*, modif*, expression), DNA near4 (transf*, modif*, recombinant), engineer* near4 (gene*, protein) [1]

---

[1] Symbols in the filter specification are defined as follows:
',': logical 'or', for example, 'livestock, cattle' means 'livestock' or 'cattle'.
'?': single character wild card, for example, 'animal?' refers to 'animal' or 'animals'.
'??': double characters wild card, for example, 'fish??' refers to 'fish' or 'fished'.

Step 3. Combine results from Steps 1 and 2 and eliminate duplicates.

Step 4. Find patents from Step 3 with agricultural keywords in titles, abstracts, or claims:

animal?, livestock, cattle, ruminant?, bovine?, ungulat*, steer?, heifer?, beef, lamb, sheep, goat?, swine, pig?, pork, poultry, avian?, chicken?, goose, geese, duck?, turkey?, fowl, fish??, agricult*, plant?, food, foodstuff?, foodcrop?, crop?, grain?, cereal?, corn, maize, wheat, barley, rye, oat?, spelt, alfalfa, vegetab*, rape, rapeseed, sorgum, millet, soy?, soybean?, legum*, bean?, tomat*, potato*, squash, onion?, leaf*, fruit?, apple?, peach??, pear?, cherr*, cotton, rice

Step 5. Identify patents from Step 4 which contain the following human drug or therapy terms, or other strong hints of being irrelevant to agricultural biotechnology:

model near5 human*, model near5 non-human, model near5 animal*, human near5 blood, human near5 hemoglob*, human near5 antibod*, human near5 antigen*, human near5 encod*, human near5 TRK*,
laboratory adj1 animal*, monkey*, rat, rats, mouse, mice, dog, dogs, cat, cats, feline*, canine*, transgenic near5 packag*, transgenic near5 non-human, animal near5 study,

---

(Continued)
'*': wild card not restricted to any particular number of characters, for example, 'agricult' refers to any words containing it, such as 'agriculture' and 'agricultural'.
'near 4': within 4 words, either direction, for example, 'gene near 4 expression' refers to 'gene' and 'expression' wherever there are at most four words between these two, such as " 'expression of gene.'

animal near5 studies, cancer*, allerg*, alzheimer*, skin near5 disorder*, insulin, diabet*, islet*

Step 6: Examine patent titles and abstracts found in Step 5 to see if any are relevant to agbiotech. The remainder are non-ag-biotech patents.

Step 7: Delete from patents found in Step 4 the non-ag-biotech patents found in Step 6.

Patents remaining after Step 7 form the basis of our final patent database.

Some believe the U.S. Patent Office Classification (USPOC) is more precise, as it is updated more frequently and has more categories at the subclass level than have the International Patent Classification (IPC) and other patent classification systems. However, as we need a classification for biotechnology that is consistent only over our study period, and the present study is conducted at a fairly aggregate classification level, the IPC is not significantly less 'precise' than is the USPOC. On the other hand, the IPC is much better organized for grouping subject matter in a hierarchical fashion and is considered by CHI to be better arranged by industry area rather than by invention art. Thus, CHI Research chooses to base its database categories on the first-given IPC (a patent is usually assigned to several IPC classes and the first one is the main assignment) to avoid misassignments of patents to industry areas (CHI Research). I originally intended to include ag-biotech patents recorded in the European Patent Office database, but later found this infeasible as CHI Research had not yet distinguished scientific from nonscientific references in EPO patents in recent years. Nevertheless, using the IPC allows us to apply the same filter consistently to both the U.S. and European Patent Office databases. It will thus enable us to extend the study to include European-registered patents when the opportunity arrives. For these reasons, I choose to use the International Patent Classification for filtering purposes in the present research.

Group C12N 15 in the IPC, entitled 'mutation or genetic engineering', 'covers processes wherein there is a modification of the genetic material which would not normally occur in nature without intervention of man and which produce a change in the gene structure which is passed on to succeeding generations.' (7th Edition of IPC). Except for subgroup C12N 15/01 dealing with mutation, all other subgroups within C12N 15, namely C12N 15/02 through C12N 15/90, refer to biotechnology. According to the IPC manual, subgroups C12N 15/02 – 08 refer to 'the preparation of hybrid cells by fusion of two or more cells', and subgroups C12N 15/09 – 90 refer to 'recombinant DNA-technology'. (For more information about IPC classes, see the 7th edition of the IPC, available online at http://classifications.wipo.int/fulltext/new_ipc/ index.htm.) While the title of the first of these two categories seems not to include gene modification technology, most patents in this category carry transgenic or recombinant DNA keywords in their titles and abstracts, as detected in Step 2. Hence, we retain subgroups C12N 15/02 – 08 of this first category in our filter.

In USPOC class 935, defined as 'genetic engineering: recombinant DNA technology, hybrid or fused cell technology, and manipulations of nucleic acids', is the explicit biotechnology class. There is significant overlap between USPOC class 935 and IPC group C12N 15, although they don't correspond to each other exactly.

In some IPC classes such as C12N 5 (undifferentiated human, animal or plant cells, tissues), some patents may be biotechnological but are not assigned into any explicit biotechnology IPC groups. In order to include those patents, we use certain keywords referring to transgenic or recombinant DNA in addition to classifications that are explicitly biotechnological.

After all the patents considered to be biotechnological were drawn, selected agricultural keywords were employed to restrict the set to those directly connected to agriculture. These keywords include broad ones such as 'agriculture', 'food', 'crop', 'animal', and 'livestock', as well as the names of specific agricultural commodities. As a check of the completeness of this keyword list, I searched patents containing the names of commodities not included in the list. For example, I checked all patents in 1998, 1999, and 2000 for those containing the term 'grape', 'grapes', 'grapevine', or 'grapevines'. Among the 76 'grape*' patents, just four involve transgenic plants or recombinant DNA; all four of these had been identified in our filter even though we had not explicitly included the term 'grape*' in it. They had been captured because they had fallen into the correct biotech IPC class, or contained the transgenic or rDNA keywords, and included the term 'plant'. This reassures us that our agricultural keyword list includes most or all innovations that have been agricultural.

Noting that the patents returned in the above steps may, however, include some that are not specific to agriculture, I next employed selected terms to exclude them. The excluded categories are: (1) transgenic plants and animals that are primarily for use in humans or on human diseases, e.g., 'human antibody', 'human antigen', and 'diabetes'; (2) any animals or plants used entirely for research purposes, e.g., 'mouse', 'rat', and 'monkey'; (3) any animals primarily designated as pets, e.g., 'dog' and 'cat'. As some of the patents so excluded might in fact be relevant to ag-biotech, a manual culling process is undertaken by reading through the excluded patents' titles and abstracts and forming a judgment about which ones are indeed non-ag-biotech. I followed this approach because a manual examination of patents assigned to certain medical and pharmaceutical classes such as A61K revealed that the majority of them are clearly agriculturally related.

Deleting the non-ag-biotech patents in this fashion, we are left with a selected set of agricultural biotechnology patents. In the final step of our search strategy, I examine the "neighbors" of those in the selected set, defined as those directly cited by or directly citing the patents in the selected set, except for those in the selected set themselves. From among these neighbors, I identify those containing any agricultural keyword included in the above-defined filter. I then examine the titles, abstracts, and classifications of the so identified patents to determine whether they should be added to the selected set. If more patents are added in this way, our filter is effectively supplemented and our final patent set is more complete. It is, at the same time, a good test of our filter's robustness. Were only a few patents so added, the filter would presumably be fairly robust and complete. Adding such patents to the selected set gives us the final set of agricultural biotechnology patents for our study.

5.1.1.2 Variables

Using the search strategy and filter specified in the previous section, I draw a set of 1,746 agricultural biotechnology patents issued between 1985 and August 2000 from the U.S. Patent Office database. An ag-biotech patent document is given for illustration in Appendix B. For each of these patents, the information on the front pages of the patent document is observed and summarized by the following variables:

(a) Patent number assigned by the Patent Office

(b) Application date

(c) Issue date

(d) Title of the patent

(e) Abstract of the patent

(f) USPOC classes assigned by the Patent Office and the position of each class in the list of USPOC classes for this patent.

(g) IPC classes assigned by the Patent Office and the position of each class in the list of IPC classes for this patent.

(h) Assignee names at the time of patent issuance, the position of each assignee in the list of assignees for this patent, and the current corporate parent names[1] of each assignee, as unified by CHI.

[1] A subsidiary owned more than 50% by a parent is counted as wholly owned by the parent. CHI keeps up-to-date track of the organizational and name changes of the numerous companies and joint ventures that are awarded patents, including parent-subsidiary relationships, mergers, divestitures, and patent re-assignments.

(i) Inventor names, the position of each inventor in the list of inventors of this patent, and each inventor's city, state, and country.

(j) Number of claims

(k) Number of patents in the U.S. and European Patent Office (EPO) databases citing this patent, and the patent number, application date, and issue date of each citing patent.[1]

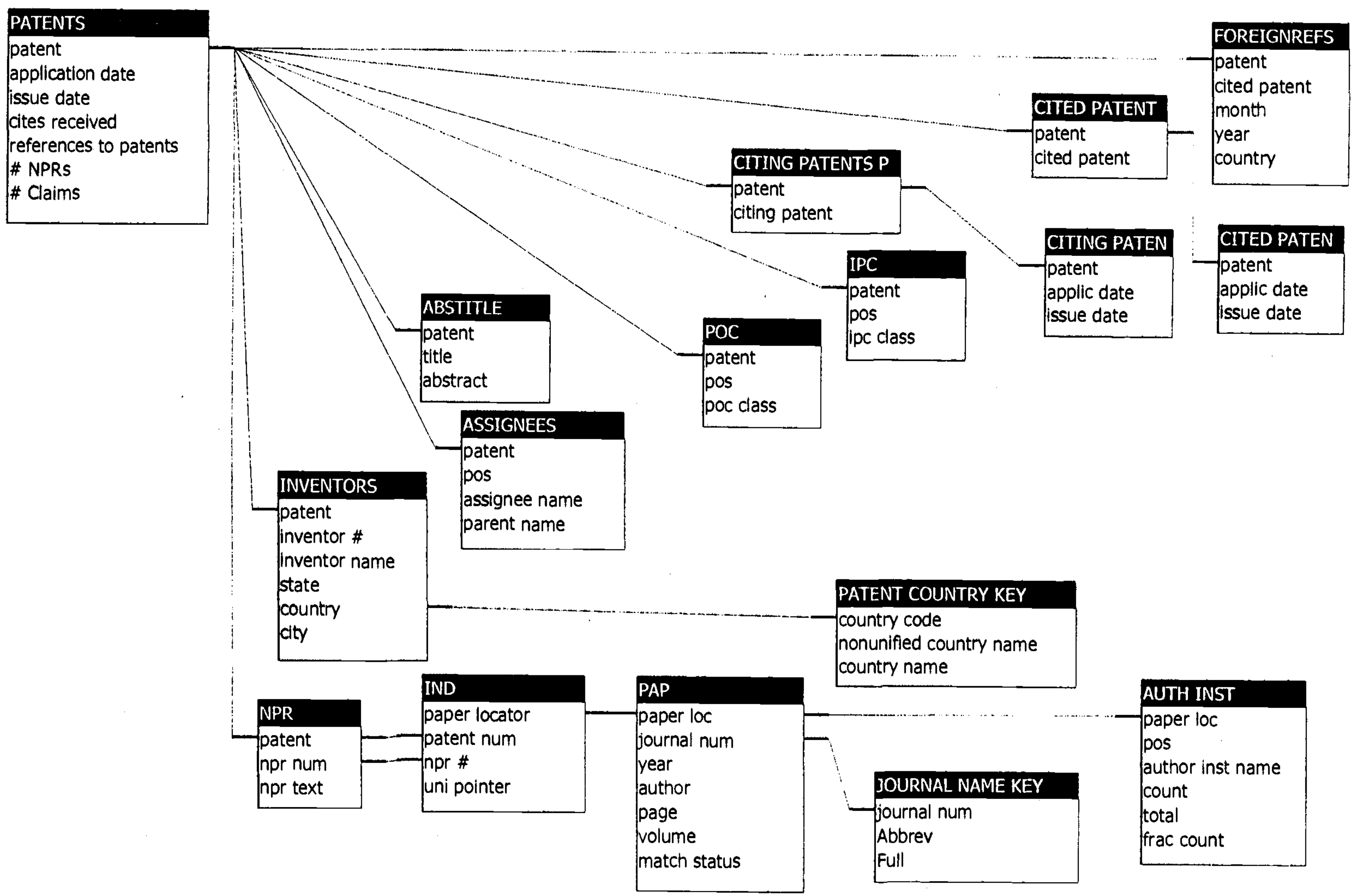
(l) Number of patents in the U.S. and EPO databases cited by this patent, and the patent number, application date, and issue date of each cited patent.

(m) Patent number, issue year, and issue country of each of the foreign (non-U.S. or EPO) patents cited by this patent.

Some of these variables, such as patent number, title, abstract, and application and issue dates, have single values on a given patent, whereas other variables, such as inventors, assignees, patent classifications, and citing and cited patents, have multiple values. For the sake of avoiding repetition of information about the same patent on a simple spreadsheet, the data on the listed variables are saved in several tables linked by the patent numbers in a relational database. The relationships among those tables are presented in figure 5.1. A description of the tables and their component data fields can be found in Appendix C.

[1] Citing patents are recorded as of January 2001.

Figure 5.1 Relationship Among Tables in the Relational Agbiotech Patent and Bioscience Paper Database



5.1.2 Scientific Papers

5.1.2.1 Identification of Scientific Papers

Apart from the information discussed in the previous section, the front pages of each patent document contain a bibliographic list of non-patent references consisting of scientific and nonscientific publications. CHI reports that approximately two million publications are referenced in the two million biotech and non-biotech U.S. patent documents that it has observed. Since one million of these references were to nonscientific material, such as to congressional hearings and newspapers, only one-half of all references have been to scientific papers — journal articles or papers presented at professional meetings. CHI Research has discriminated between the scientific and the nonscientific references. For example, among the ten non-patent references listed on the front pages of a certain patent, CHI can identify eight as scientific papers, six of which have full bibliographic information. If only five of those six are found in the library, these five are identified as scientific papers cited by the given patent.

5.1.2.2 Variables

In our selected set of 1,746 agricultural biotechnology patents, scientific and nonscientific publications are cited 43,918 times. Each patent is linked to scientific papers cited by this patent by pairing the patent number and the paper locators assigned by CHI. [1] A total of 30,792 of such pairings represent all linkages that can be verified between the 1,746 ag-biotech patents and the 13,325 scientific papers cited by such patents in a one-to-many correspondence. It should be noted that one paper can also be cited by more than one patent. Consequently, the linkage from a paper to its citing patents is also a one-to-many relationship.

For each scientific paper, we have observed the following information:

(a) Paper locator

(b) Journal name abbreviation

(c) Full journal name

(d) Publication year

(e) Volume of journal

(f) Page

(g) Abbreviated names of authors

CHI Research then matches these data to the Institute of Science Information's (ISI) Science Citation Index (SCI) to identify the institutions at which the scientists worked at time of publication. SCI lists only the addresses of authors' affiliations at the department or school level, not at the level of the home institution of each separate author. This creates problems in assigning weights to different author institutions when multiple authorships are listed on a paper.

Consider, for example, a case in which a bioscience article cited by an agricultural biotech patent is authored by three scientists, the first and third being

---

[1] CHI Research assigns a unique paper locator to each paper.

affiliated with the Department of Biochemistry at Michigan State University and the second with the Department of Molecular Biology and Medical School at Cornell University. If we give weight to each author institution based on the proportion of the number of authors associated with that institution to the total number of authors in the paper, a weight of 2/3 should be assigned to Michigan State University and 1/3 to Cornell University. In the Science Citation Index, however, the paper's author affiliations will be listed as three addresses: Department of Biochemistry at Michigan State University, Department of Molecular Biology at Cornell University, and Medical School at Cornell University. Hence, under a weighting system based on the mentions of each institution in the list of authors' affiliation addresses, Michigan State University and Cornell University will instead be given 1/3 and 2/3 credits for this paper, respectively. Although the first weighting method reflects more precisely the contribution of each institution to the given paper, we can follow only the second method, as SCI is the solely complete database providing information on author institutions in scientific literature.

CHI Research develops two approaches for assigning credits to each author institution. In the first approach, a fraction of one credit is assigned to an author institution of a given paper, as discussed previously. The credit assigned this way is called the "fractional count." In the other approach, one credit, or a "whole count," is given to each author institution in a given paper. Using this approach, Michigan State University and Cornell University will receive one credit each in the above example.

It was found that among the 13,325 scientific papers cited by our selected set of 1,746 agricultural biotechnology patents, a subset of 8,099 were from at least one U.S. institution and published between 1973 and 1997. The cited papers published in 1998, 1999, and 2000 were not included since when CHI prepared the database for this study, it had not yet cleaned the citations to those papers in patents issued in 2000. However, only a very few papers were missed by this process considering the lag between the publication date of a cited paper and the application date of the citing patent, and the lag between the patent's application and issue dates, which usually are more than 18 months and a year, respectively. For example, in patents issued in 1999 in our dataset, no citations were to papers published in 1998 or 1999, and less than 1% of the citations were to papers published in 1997. Similarly, we expect the number of citations from patents issued in 2000 to papers published in 1999 and 2000 to be extremely low or zero, and citations to papers published in 1998 to account for less than 1% of total paper citations.

In each of the 8,099 scientific papers with at least one U.S. author institution, the following variables have been identified:

(a) Paper locator

(b) Author institution names

(c) Position of each institution in the list of author institution names

(d) Total count of author institutions' addresses at department or school levels

(e) Number of addresses associated with each institution for this paper

(f) Fractional count of each institution ($f = e / d <= 1$)

Data on the paper side were saved in several tables linked by the paper locators. These paper-related tables were then linked with the patent-related tables discussed in section 5.1.2 by pairing patent numbers and paper locators in an index table. Summaries of the paper-related tables and relationships among the various patent-related and paper-related tables are illustrated in figure 5.1. Description of those tables and of their component data fields is given in Appendix C.

5.2 Research Inputs

Data on research input allocations and other characteristics of the identified patent assignees and paper author institutions come from a number of sources, including the National Science Foundation (NSF), the National Center for Education Statistics (NCES) of the U.S. Department of Education, the Association of University Technology Managers (AUTM), the National Research Council (NRC), the Gourman Report of Graduate Programs, the National Association of State Universities and Land Grant Colleges, and the Bureau of Census. Patent assignees and scientific paper author institutions identified in our patent and paper datasets, respectively, fall into five sectors: universities and colleges, Federally Financed Research and Development Centers (FFRDCs), nonprofit institutions, government agencies, and private firms, all except the last of which are not profit-making organizations. Data sources and data fields differ by sector, and the detailed data discussion below refers only to the university and private firm sectors. Descriptions of data in the other three sectors are given in Appendix D.

5.2.1 Research Inputs at Universities and Colleges

Universities and colleges are institutions of higher education engaged primarily in providing at least 2 years of college-level study leading toward a degree. In this section, university and college data are discussed according to source.

5.2.1.1 National Science Foundation

Data on research inputs at universities and colleges are, for the most part, available from the WebCASPAR database of the National Science Foundation. They include information by individual university or college, as obtained from a number of surveys described below. The first four of these surveys are conducted by NSF and the last by the NCES of the U.S. Department of Education.

The Survey of Research and Development Expenditures at Universities and Colleges (SRDEUC) is the primary source of information on separately budgeted R&D expenditures within U.S. academia. It has been conducted, annually since fiscal year 1972, in the target population of institutions that have doctoral or master's programs in the sciences or engineering and that perform at least $150,000 worth of separately budgeted R&D each year. The FY 1978 survey employed a different population and different questions than in preceding or subsequent surveys, so that information in that year lacks comparability with other years. However, FY 1978 data still will be included in our empirical analysis in order to render the time series complete. The key variables reported in SRDEUC include academic institution, R&D expenditures, research equipment expenditures, field of science and engineering, sources of funds (federal, state and local, industry, institutional, and other) and capital expenditures.

The Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions is the only comprehensive source of data on federal science and engineering funding to individual academic and nonprofit institutions. It has been conducted annually since 1965 and was expanded in 1971 to include information on science and engineering (S&E) fields. Universities and colleges have been covered in every survey year, while coverage of other sectors, such as FFRDCs and non-profit organizations, in this survey have changed. The target population consists of as many as 21 federal agencies, which together incur virtually all obligations for federally financed academic R&D. Key variables included in the survey are academic institution, federal agency, obligation for S&E by obligating agency, performer of the S&E work, and type of activity (e.g., R&D, R&D plant, and S&E instructional facilities).

The Survey of Federal Funds for Research and Development provides information about federal funding for R&D in the United States. It has been conducted annually since the early 1950's, and the information included in this survey has been fairly stable since 1973. The survey's target population consists of all federal agencies, and in some cases some subdivisions of agencies, which conduct research and development programs. The following key variables are included: federal agency, federal obligations and outlays for R&D, field of science and engineering, performer (type of organization doing work), and character of work (basic research, applied research, and development).

The Survey of Graduate Students and Postdoctorates in Science and Engineering collects data on the number and characteristics of graduate students and postdoctorates in all institutions offering graduate degree programs in a science or engineering field, including branch campuses, affiliated research centers, and separately organized components such as medical schools. This annual survey was first conducted in 1966 among a limited number of doctorate-granting institutions, and was expanded since 1975 to include all institutions known to have programs leading to the master's or doctorate degrees in science and engineering. Although NSF has attempted to maintain consistent trend data, some modifications have been made in the data collection to respond to changing issues over the past 25 years. Key variables reported in this survey are academic institution, number of graduate students and postdoctorates, field of study, mechanism of financial support (e.g., fellowship, research assistantship, etc.), primary source of financial support (e.g., NSF, NIH, USDA, etc.), enrollment status (full-time versus part-time), citizenship, and sex.

The Integrated Postsecondary Education Data System (IPEDS) is an integrated system of surveys conducted by the NCES to collect information on the faculty, students, and characteristics of all primary providers of postsecondary education. Prior to IPEDS, some of the same information had been collected in the Higher Education General Information Survey (HEGIS). These surveys have been conducted annually since 1965 and are targeted at all accredited 2- and 4-year postsecondary educational institutions in the United States. Key variables include academic institution, faculty numbers and average salary, master's and doctoral degrees awarded, enrollment, tuition, and financial statistics.

Some of the data in the WebCASPAR database are available only by individual university or college, while other data are further broken down into science and engineering fields. Agricultural biotechnology research and bioscience research facilitating agricultural biotech research fall into three fields in WebCASPAR: agricultural, biological, and medical sciences, all listed under the life science category. However, biotechnology research constitutes only part of these three fields. For example, landscape architecture in agricultural science and public health in medical science do not include biotechnology-related research. Given the disaggregated level of our data on research outputs, we ideally would like to match input and output data such that data on inputs to research projects correspond to the outputs produced by those same inputs. However, input data divided into agricultural, biological, and medical research are the most disaggregate form readily available to us. Under the limitations of our data sources, these data provide the best approximations to research input measures at the institutions of interest in our study. Other data at the institution level can be used to capture the fixed factors and institutional characteristics.

The NSF's WebCASPAR database provides annual data on the following variables, broken down by S&E field at individual universities or colleges:

(a) Total R&D expenditures

(b) Federally financed R&D expenditures

(c) Current total research equipment expenditures

(d) Current federally financed research equipment expenditures

(e) Total capital expenditures

(f) Federally financed capital expenditures

(g) Number of graduate students

(h) Number of postdoctorates

(i) Number of Master's degrees awarded

(j) Number of Ph.D. degrees awarded

Variables (a) and (b) are available annually from 1973 to 1998, (c) and (d) annually from 1981 to 1999, (e) and (f) annually from 1972 to 1989, and (g) through (j) annually from 1972 to 1998.

WebCASPAR includes annual data on the following variables at the institutional (individual university or college) level only:

(a) R&D expenditures by financing source (e.g., federal government, state and local government, industry, institution, and others)

(b) Federal obligations [1] for S&E by type of activity (e.g., R&D, R&D plant, instruction facilities, fellowships, general support for S&E, and other S&E support) from each obligating federal government agency

---

[1] Obligations are the amounts for orders placed, contracts awarded, services received, and similar transactions during a given period, regardless of when the funds were appropriated and when future payment of money is required. Obligations differ from expenditures in that funds allocated by federal agencies during one fiscal year may be spent by the recipient institution either partially or entirely during one or more subsequent years.

(c) Number of faculty members by academic rank (e.g., full professor, associate professor, assistant professor, lecturer, and instructor)

(d) Number of tenured faculty by academic rank

(e) Total faculty salary outlay and average faculty salary by academic rank

(f) Opening fall enrollment by enrollment level (undergraduate, graduate, first professional) and enrollment status (full-time versus part-time)

(g) Tuition by enrollment level and tuition type (e.g., in-state, out-state, and in-district)

Items (a), (b) through (e), and (f) through (g) are available from 1973 to 1998, 1971 to 1998, and 1967 to 1997, respectively.

Apart from data by individual university or college in the WebCASPAR database, other data aggregated across universities and colleges are provided on an annual basis in some NSF report series, such as in "Academic Research and Development Expenditures" and "Federal Funds for Research and Development." These data items include:

(a) Total R&D expenditures by character of work (i.e., basic research, applied research, and development [1])

(b) Federally financed R&D expenditures by character of work

---

[1] NSF defines research as systematic study directed toward fuller knowledge or understanding of the subject studied. Research is classified as either basic or applied, according to the objectives of the investigator. Basic research is directed toward an increase of knowledge only, whereas in the applied research, the primary aim of the investigator is a specific application of the new knowledge. Development refers to, according to the NSF, systematic use of the knowledge or understanding gained from research, directed toward the production of useful materials, devices, systems, or methods, including design and development of prototypes and processes.

(c) Federal obligations for total research, basic research, and applied research by field of science and engineering (e.g., agricultural, biological, and medical sciences) for each obligating federal government agency (e.g., USDA, NIH, NSF, etc.)

These three data items are reported for the periods 1953 to 1998, 1972 to 1998, and 1973 to 2000, respectively. They can be used in conjunction with WebCASPAR data to compute the proportions of basic and applied research in total R&D expenditures, or federal obligations in the life sciences at individual universities or colleges.

5.2.1.2 National Center for Education Statistics, U.S. Department of Education

The following information on an individual university's location and characteristics is available from the Institutional Characteristics Data Files in the Integrated Postsecondary Education Data System (IPEDS), which is maintained by the NCES of the U.S. Department of Education:

(a) Institution name
(b) Post office state abbreviation code
(c) County name
(d) City location of institution
(e) Institution street address
(f) Zip code

(g) Bureau of Economic Analysis (BEA) region code (e.g., 01 – New England, 02 – Mid East, 03 – Great Lakes, 04 – Plains, 05 – Southeast, 06 – Southwest, 07 – Rocky Mountains, 08 – Far West, and 09 – Outlying Areas)

(h) Degree of Urbanization (i.e., large city, mid-size city, urban fringe of large city, urban fringe of mid-size city, large town, small town, and rural area designated by the Census Bureau)

(i) Control of institution (e.g., public, private nonprofit, and private for-profit)

(j) Carnegie Foundation classification (e.g., research university, doctoral university, master's university, specialized institution.)

(k) Highest degree offered

(l) Presence or absence of hospital

(m) Whether or not a medical degree

These variables do not change much over time and can be used to develop measures of institutional fixed factors.

5.2.1.3 Association of University Technology Managers

The Association of University Technology Managers' (AUTM) Annual Licensing Survey is the only one that gathers technology licensing data from academic institutions. It has been conducted since 1991. The technology transfer infrastructure at universities and colleges can be captured by some variables in this survey, reported on an annual basis and at the institution level. The variables include:

(a) Licensing FTEs in technology transfer office
(b) Licenses and options executed by licensee (i.e., start-up, small company, and large company)
(c) Total active licenses and options, by field (e.g., life science and physical science)
(d) Gross license income received, by field
(e) Invention disclosures received
(f) Total and new U.S. patent applications filed
(g) U.S. patents issued
(h) Start-up companies formed

The data are available annually from 1991 to 1999 on a sample of 130 to 180 institutions. The sample frame changes slightly across years but continues to cover 85% of the top 100 U.S. universities as ranked by research dollar volume.

5.2.1.4 Other Sources

The National Research Council (NRC) of the National Academy of Science conducted two studies (1982 and 1993) to assess the status and quality of research-doctorate programs in the sciences (including biological sciences, physical sciences and mathematics, and social and behavioral sciences), in engineering, and in arts and humanities in the United States. The 1993 study reported information on 3,634 programs in 41 fields at 274 universities. This sample size was about 35% greater than that included in the 1982 study. Data were collected from a number of sources, including the Institution Coordinator Response Data, the Institute for Science Information's (ISI) Science Citation Index (SCI), and the National Survey of Graduate Faculty, in which a sample of faculty was asked to comment on the quality of programs in their own fields. The fields of interest in the present research are those in the biological sciences, namely biochemistry and molecular biology, cell and developmental biology, and molecular and general genetics. The following information is available on these three fields at individual participating universities:

(a) Scholarly quality of program faculty (five-point scale rating, zero signifying "not sufficient for doctoral education" and 5 signifying "distinguished")
(b) Program effectiveness in educating research scholars and scientists (five-point scale rating, zero representing "not effective" and 5 representing "extremely effective")
(c) Program ranking
(d) Total number of faculty in the program
(e) Percentage of faculty with federal support
(f) Publications per faculty
(g) Citations per faculty

Items (a) through (e) were collected in the two study years, and items (f) and (g) were summed over the four-year periods prior to the study years, that is, over periods 1977 to 1981 and 1988 to 1992.

Another source of graduate program ranking is *The Gourman Report*, which has been updated in 1985, 1987, 1993, and 1995, more frequently than the NRC's report. Both agricultural and biological sciences are included in the Gourman guide, but ratings are given only to the programs in the top-quality group in each field, i.e., those with a rank of 4 or above on a scale of 1 to 5. Using this value for a program with 4 or above ranking and zero for a program ranked below 4 gives a positive rank to the first 30-50 programs in each field and then rates all the others as equally poor. The quality of agricultural programs at a university can be measured by summing the ratings of this university's graduate programs in the fields of agricultural sciences, such as animal science, horticulture, botany, and entomology. Similarly, summing the ratings of graduate programs in the fields of biological sciences including biochemistry, cellular and molecular biology, genetics, and microbiology gives the quality of biological programs at that university.

The National Association of State Universities and Land Grant Colleges (NASULGC) provides a list of land grant universities. Matching the universities in our study with the ones in that list, we can decide whether a university is land grant or not.

A university's research infrastructure may affect its research productivity. The interdisciplinary biotechnology research centers at some universities may enhance biotech-related research by bringing scientists from different disciplines together and facilitating communication among them. Whether such a research exists at a particular university can be determined at the university's website.

5.2.2 Research Inputs at Private Firms

Private firms are organizations that may legally distribute net earnings to individuals or other organizations. The primary source of information on R&D performed by industry within the U.S. is the Survey of Industry Research and Development (Form RD-1), sponsored by the NSF and enumerated by the Census Bureau's Manufacturing and Construction Division. It is an annual sample survey conducted since the early 1950s and representing all for-profit R&D-performing companies in the nonfarm industries, either publicly or privately held. The Standard Statistical Establishment List (SSEL), a Census Bureau's compilation containing information on more than three million establishments[1] with paid employees, is the target population from which the frame used to select the survey sample is created. For companies with more than one establishment, data are summed to the company level. Each company is assigned a single Standard Industrial Classification (SIC) code at the 3-digit level based on the activity of the establishment with the highest dollar value of payroll. Some large companies are surveyed annually. Before 1994, they consisted of all firms with 1,000 or more employees. This criterion was changed to $1 million in R&D expenditure in the 1995 survey and to $5 million in R&D expenditure

---

[1] An establishment is defined as an economic unit, at a single location, where business is conducted or where services or industrial operations are performed.

in the 1996 survey. Small companies have been surveyed at irregular intervals: in 1967, 1971, 1976, 1981, 1987, 1992, and 1996. Although since the survey was first fielded, some improvements have been made in sampling, collection, processing, and tabulation methods, the survey has covered more than 90% of industrial R&D and provided good indicators of industry R&D spending and personnel.

Aggregate statistics from the industrial R&D survey are published in the annual NSF report, "Research and Development in Industry at the 2- and 3-digit SIC Industry Levels." Firm-specific information is held confidentially because of potential disclosure of information about particular respondents, who are guaranteed anonymity by law. Fortunately, data files with the microdata are maintained by the Census Bureau's Center for Economic Studies (CES) and are housed at the Research Data Center at the University of California, Berkeley, and other sites.

I accessed, for individual agricultural biotech firms, the following RD-1 information for years 1973 to 2000 from the CES, Bureau of Census microdata files:

(a) Total R&D expenditures

(b) Federally financed R&D expenditures

(c) Company financed R&D expenditures

(d) Distribution of R&D by state

(e) Distribution of total R&D between basic research, applied research, and development

(f) Total pollution abatement R&D

(g) Employment of R&D scientists and engineers

(h) Total employment

(i) Sales volume

(j) Firm's geographic location

(k) Standard Industrial Classification (SIC) code

Items (a), (b), (i), and (j) are available annually from the large surveyed firms, and in survey years from small firms. In non-survey years, the latter data may be imputed. The remaining items are provided voluntarily by large firms in odd-numbered years (except annually from 1973 to 1977) and contain a number of missing values.

Most firms included in our study are engaged in more than agricultural biotechnology R&D and agricultural production. For example, in its annual 10-K form filed with the Security and Exchange Commission (SEC), Monsanto Co. reported net sales of $3.470, $4.264, and $5.102 billion in agricultural products, and $2.323, $2.771, and $3.920 billion in pharmaceutical products in years 1997, 1998, and 1999. For companies with segment sales data available, R&D expenditures in each segment can be approximated by allocating total R&D expenditures to each segment according to the proportionate share of that segment in total sales. Using this approach, of the $1.290 billion in total R&D expenditures at Monsanto in 1999, an estimated $729.5 million were allocated to agricultural R&D and $520.5 million to pharmaceutical R&D.

This study could be extended to include conventional production function relationships in which biotech firms' patents are used as intermediate inputs.

Although I do not do so in the present dissertation, I do provide descriptions of the production outputs, inputs, and other characteristics at biotech firms identified in our patent and paper datasets. The descriptions are given in Appendix E.

Chapter 6: Results

Using the data discussed in Chapter 5 on agricultural biotechnology patents, bioscience papers, and firms' and universities' research inputs and characteristics, I now estimate the models specified in Chapter 4. I then use these estimates to characterize the production possibility frontiers (PPFs) between the multiple outputs in the university and firm models. Measures of returns to R&D are computed under alternative policy settings and elasticities derived. Finally, tests are developed of the complementarity or substitutability between basic bioscience and applied biotechnology research.

6.1 Results of University Model Estimation

6.1.1 Parameter Estimates

The university model consists of the two equations (4.8) and (4.9), which respectively explain the production of patent-cited bioscience articles and graduate education. The model is estimated using the balanced, pooled, 1985-1997 time-series cross-sectional data on 177 universities, as described earlier.

As discussed in section 4.2.1, university characteristics variables such as *Faculty*, *Enroll*, *FacSalary*, *LifeShare*, *ResearchIntense*, and *TeachIntense*, and fixed factors such as *LandGrant*, *MedSchool*, *PrivPub*, and *BEARegion* , were initially included in the estimation stepwise. However, the inclusion of these variables worsens multicollinearity, extracts explanatory power from the R&D variables, and brings little improvement in the goodness-of-fit. The fact that most of the characteristics and fixed factors are nonsignificant implies that, after controlling for university R&D expenditures, numbers of post-docs, and graduate program rankings in agricultural and biological sciences, a university's characteristics and fixed factors such as its overall size, overall faculty quality, orientation toward life science, research and teaching intensity, land grant status, medical school presence, public vs. private school status, and geographical location do not significantly affect the university's efficacy in producing agbiotech-cited scientific papers or in providing graduate education in the agricultural and biological sciences. Hence, university characteristics and fixed factors are excluded from the regressions reported below. The simple correlation coefficient between aggregate private agricultural R&D expenditures (*PrivRD*) and a time trend was 0.95. Of these two, only *PrivRD* was therefore left in the two equations.

The annual number of agbiotech patents awarded to universities, denoted by *UnivAgPt*, constitutes our measure of university applied research output. It was first included in the university model as the third output, in exactly the same manner as were the other two outputs, *Science* and *Grad.* Since the majority of agbiotech patents in our database were issued after 1997, only 141 of them were awarded to the 177 universities during the study period of 1985 to 1997. The goodness-of-fit of the regression relating universities' agbiotech patents to their R&D expenditures and other inputs is very poor (R^2 below 0.1). To simplify the estimation and construction of the university's PPF, only a simple linear version of *UnivAgPt* is included in the two estimation equations. *UnivAgPt* provides some idea about the relationships between basic and applied research and between graduate education and applied research in a university environment.

The two equations were estimated in a number of alternative ways. First, they were estimated separately with (a) OLS, (b) a fixed effects model, (c) a random effects model, and (d) models in which a heteroskedasticity correction was made to the OLS and fixed effects estimates. Then the equations were estimated jointly with SUR. Parameter estimates, corresponding t-statistics, and the goodness-of-fit of the two equations, using each individual estimation technique, are presented in tables 6.1 and 6.2.

In the single equation estimations, the R^2s of the paper and graduate student regressions range from 0.53 to 0.65, and from 0.74 to 0.98, respectively[1]. The system weighted R^2 in SUR estimation is 0.99. These reasonably high R^2s indicate a fairly close fit of the two equations to the data. In most models, marginal products of all input are positive.

In both the science and graduate student equations, OLS and SUR produce rather similar coefficient estimates, and SUR improves estimation efficiency only a little. It seems, then, that no strong contemporaneous correlation exists between the two equations' error terms, and SUR probably does not provide estimates superior to those in OLS.

[1] The R^2 of the random effects model of the graduate student regression was only 0.32, implying a random effects specification fits the data poorly.

Table 6.1. Cited Bioscience Production in U.S. Universities: Parameter Estimates

	OLS		OLS Hetero corrected		Fixed Effects		Fixed Effects Hetero corrected		Random Effects		SUR	
Variable	Estimate	t	Estimate	t	Estimate	t	Estimate	t	Estimate	t	Estimate	t
Intercept	-0.670	-0.49	0.032	0.04	-7.436	-1.96**	-7.155	-2.63**	-0.976	-0.65	-0.315	-0.23
Science $_{t-1}$	0.627	28.65**	0.531	24.37**	0.420	16.41**	0.326	13.08**	0.558	24.39**	0.627	28.63**
Grad	0.002	0.90	0.003	1.83*	-0.008	-1.15	0.000	-0.08	0.002	0.83	0.001	0.39
Grad $^{0.5}$	-0.099	-1.42	-0.089	-2.40**	-0.052	-0.24	-0.045	-0.50	-0.121	-1.41	-0.098	-1.40
AgRD	0.436	3.28**	0.193	1.65*	0.901	2.60**	0.815	2.19**	0.550	3.47**	0.490	3.68**
BioRD	0.323	3.07**	0.185	2.26**	0.108	4.20E-01	0.375	2.15**	0.310	2.44**	0.370	3.48**
AgPD	0.137	4.91**	0.203	5.72**	0.135	3.37**	0.172	3.36**	0.146	4.72**	0.145	5.2**
BioPD	0.013	5.72**	0.017	7.76**	0.035	5.73**	0.032	6.40**	0.016	6.00**	0.013	5.97**
AgRank	-0.001	-0.20	0.004	0.87	0.001	0.05	0.000	0.01	-0.003	-0.38	-0.003	-0.54
BioRank	-0.008	-1.86*	-0.007	-2.67**	-0.016	-1.77*	-0.004	-0.65	-0.010	-2**	-0.008	-1.92*
PrivAgRD	0.007	2.13**	0.003	1.86*	0.016	4.30**	0.004	2.45**	0.009	2.98**	0.007	2.11**
UnivAgPt	-0.647	-1.77*	1.040	1.43	-0.803	-2.01**	0.843	1.18	-0.664	-1.77*	-0.673	-1.84*
R^2	0.63		0.54		0.65		0.57		0.53		0.99	

Note: ** and * indicate parameters significant at 95% and 90% confidence levels, respectively.

Table 6.2. Graduate Education Production in U.S. Universities: Parameter Estimates

	OLS		OLS Hetero corrected		Fixed Effects		Fixed Effects Hetero corrected		Random Effects		SUR	
Variable	Estimate	t	Estimate	t	Estimate	t	Estimate	t	Estimate	t	Estimate	t
Intercept	313.384	11.62**	206.512	9.27**	174.358	7.75**	168.673	7.13**	226.900	12.35**	311.280	11.54**
Science	-2.759	-3.93**	-3.838	-3.88**	-0.982	-3.82**	-0.492	-1.48	-1.210	-4.33**	-3.087	-4.4**
Science $^{0.5}$	15.632	3.79**	16.601	3.72**	3.651	2.33**	1.402	0.86	5.120	3**	15.520	3.77**
AgRD	48.100	18.42**	52.300	17.63**	-4.981	-2.13**	-9.248	-3**	11.000	4.84**	48.000	18.52**
BioRD	39.300	19.05**	40.100	16.5**	9.505	5.54**	8.900	4.93**	13.000	7.1**	39.000	19.13**
AgPD	7.677	14.38**	7.264	9.68**	1.222	4.61**	0.932	2.58**	1.820	6.34**	7.800	14.6**
BioPD	0.500	11.52**	0.783	15.55**	0.519	13.37**	0.621	14.71**	0.570	14.02**	0.510	11.8**
AgRank	-1.687	-15.92**	-1.412	-13.13**	0.183	1.48	0.075	0.6	-0.765	-6.31**	-1.690	-15.93**
BioRank	-0.150	-1.71*	0.139	1.55	-0.173	-2.81**	-0.117	-1.91*	-0.193	-2.94**	-0.160	-1.78*
PrivAgRD	-0.033	-0.51	0.065	1.53	0.319	13.73**	0.298	15.21**	0.239	9.6**	-0.025	-0.38
UnivAgPt	-23.448	-3.12**	-21.967	-2.61**	-14.848	-5.61**	-12.458	-3.75**	-16.130	-5.59*	-22.830	-3.04*
R^2	0.81		0.74		0.98		0.97		0.32		0.99	

Note: ** and * indicate parameters significant at 95% and 90% confidence levels, respectively.

A Hausman test suggests no significant misspecification in the random effects model, and a joint F-test cannot reject the fixed effects model in either the paper or graduate student equations. However, comparisons of parameters from the OLS, fixed effects, and random effects models suggest it is reasonable to base our policy analysis on the OLS estimates. The random effects estimation of the paper equations are quite close to the OLS estimates, and while the fixed effects estimates are less close, they still are not far from the OLS estimates. Furthermore, the addition of the university-specific dummy variables represented by the fixed effects model raises the R^2 in the OLS specification by only 0.02, and increases also the coefficient sampling variance of such policy-relevant variables as biological research expenditures (*BioRD)*. In the graduate student equation, parameter estimates are less robust to estimation choice than they are in the paper equation. However, the random effects model provides a particularly low R^2 of 0.32 and the fixed effects model, while offering a much higher R^2, again increases the coefficient sampling variance of important input variables. Moreover, the characteristics dummy variables representing unobserved, output-relevant variations among universities were originally included in the model but were largely nonsignificant, arguing against the modeling of university-specific effects.

Heteorskedasticity was detected in the OLS estimation of the paper and graduate student equations. Unfortunately, regressions fitting the estimated error terms against the explanatory variables did not provide a very good fit and no effective correction could be developed. Indeed, as shown in tables 6.1 and 6.2, the heteroskedasticity correction attempted with the data available hardly changed the

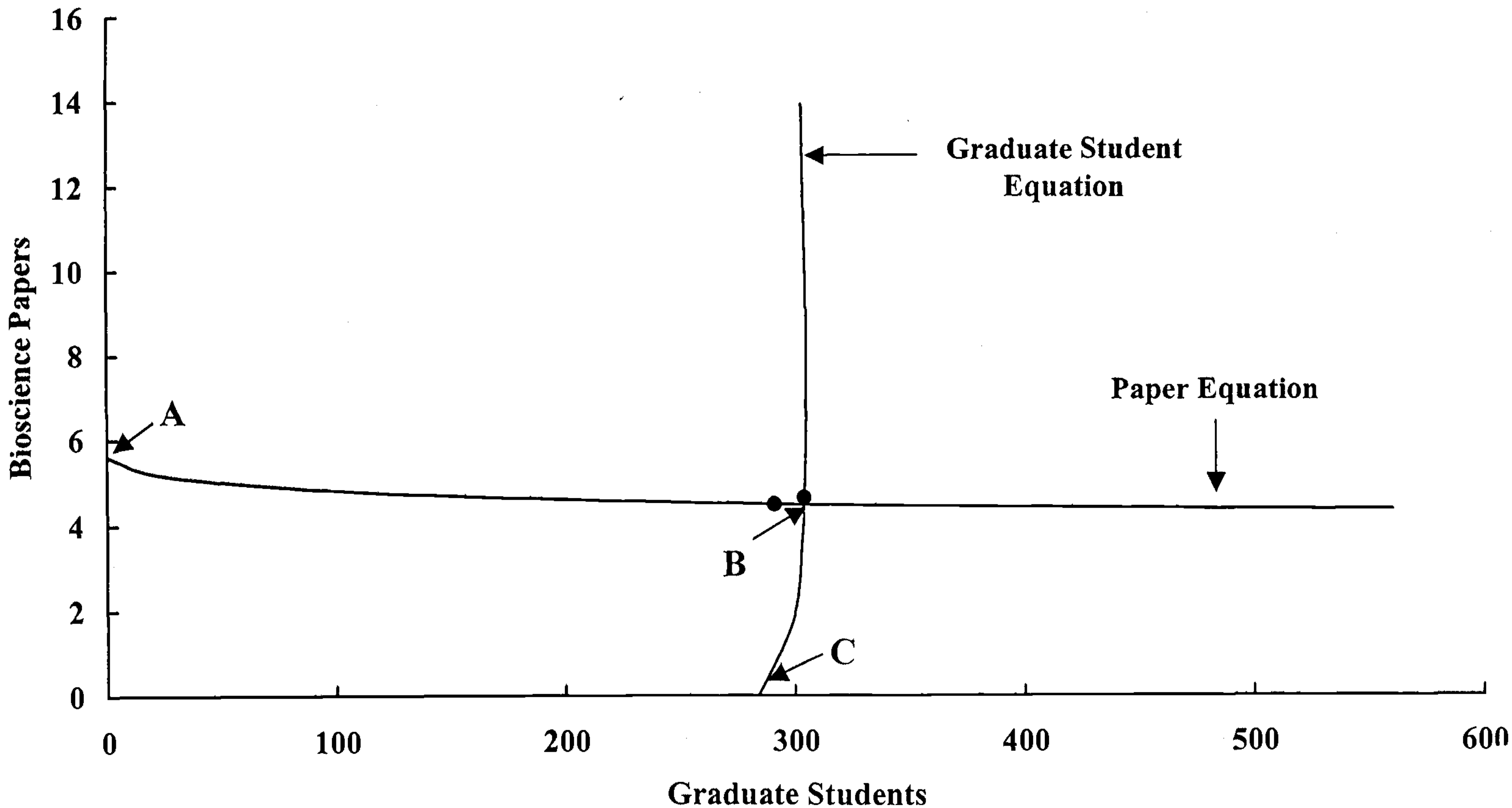
OLS parameter estimates. In the presence of heteorskedasticity, OLS estimates are, although not efficient, consistent, an important property considering the large sample used in the present study. OLS estimation produces largely significant parameter estimates. Linear and square-root expressions of graduate student numbers (*Grad*) are nonsignificant in the paper equation, but the lagged dependent variable and all inputs except *AgRank* are significant at the 10% level. In the graduate student equation, all variables except *PrivAgRD* are significant at the 10% level. In the following, I discuss the university model results based on the OLS estimation.

6.1.2 Production Possibility Frontier

Two portions of a production possibility frontier, based respectively on the science and graduate student equations in the university model, are traced in figure 6.1. The right-hand side outputs, namely the number of graduate students in the science equation and the number of patent-cited bioscience papers in the graduate student equation, are varied in figure 6.1 from zero to one standard deviation above the sample mean, and all inputs are held fixed at sample means.

Because of the right-skewed nature of the university sample, especially in the bioscience paper dimension, the bulk of the universities lies near either point A or point C. For example, the sample median and maximum of the number of cited bioscience papers published are 0.54 and 97.2, respectively; and in 74.2% of the observations, the published paper quantity is less than the 4.64 sample mean. The sample median and maximum of the number of graduate students are 207 and 1445, respectively, and in 64.5% of the observations, graduate student numbers are fewer

Figure 6.1. Bioscience Paper and Graduate Student Output in U.S. Universities: The Production Possibility Frontier



than the 297 mean value. Thus, an accurate indication of the shape of the PPF is obtained only by evaluating the paper equation at rather low graduate student numbers, and the graduate student equation at a rather low bioscience output. If instead the equations are evaluated at high levels of the opposing outputs, they will pass through a sample region in which few or no observations are found, and consequently be poorly indicative of PPF position or shape.

As discussed in section 4.6, a generalized PPF shape cannot be estimated by a single estimable regression. Alternatively, the joint set of the two PPFs based on two separate regressions can approximate a single "grand-technology" of universities' production of bioscience papers and graduate students. In figure 6.1, the portions from point A to point B and from point C to point B comprise the grand-PPF of the university model, representing various combinations of the two outputs that a science- or graduate-specializing university, respectively, can produce using the sample-mean input levels. Because this single PPF is constructed from two separate regressions and because our square-root functional form is not effective at approximating the negatively sloped portion of the generalized PPF shape in the vicinity of the sample means, a kink occurs at the intersection of the two separate PPF sections. Hence, our estimate of the grand-PPF in the vicinity of point B is rather inaccurate. However, since the square-root functional form provides its greatest curvature at lower outputs, where the bulk of universities operate and where most of the curvature might occur, our estimate of the single PPF well below point B is fairly reliable.

The portion of the grand-PPF between points A and B best reflects the technology of universities specializing in science. At this end of the PPF, the ratio of science to graduate output is high. Starting from point A, one can imagine reallocating the sample-mean input levels at a science-oriented university from science production to graduate students. Early in this reallocation, the PPF is negatively sloped and the two outputs are strong substitutes: graduate student production rises as science output drops. At point A, increasing graduate student output from zero to one leads to a reduction of 0.097 in science output. Increasing graduate student output by an additional unit brings a 0.047 increase in science output. As reallocation proceeds, this negatively sloped PPF becomes flatter. Graduate education is still a substitute for science research, but at a lower rate. Eventually, education becomes rather supplementary to science. For example, if the representative university is training 150 graduate students, the cost in cited publication output brought about by adding one more graduate student is only 0.002.

If a university focuses only on generating cited bioscience papers and does not provide any graduate education, it can produce 5.60 bioscience papers using the sample-mean input quantities. As it begins to produce graduate students, fixed costs are incurred such as lab space, computing facilities, and faculty advisors. These fixed costs, as well as associated variable costs, must be compensated by a reduction in expenditures that would be allocated to cited science. At low graduate student numbers, the effect of this reduction on science output far exceeds the services these graduate students provide toward science output, so that science output decreases. As the graduate student number rises, however, the marginal cost of another graduate student initially drops. The service this additional student provides to science output comes closer to the loss in science production induced by the reallocation of money from science to graduate students. When the two effects on science cancel out, the slope of the PPF will be zero, and graduate students will be perfect supplements to bioscience papers.

Similarly, the portion between points C and B on the single PPF in figure 6.1 reflects the technology of teaching universities, where graduate student production is the main goal and the ratio of cited publications to graduate students is low. Starting from point C, the sample-mean fixed inputs are reallocated from graduate education to bioscience. Between C and B, both graduate numbers and bioscience rise because the latter is an input to the former. At point C, raising cited publication output from zero to one per year increases the number of graduate students by 13. Further input reallocation to science flattens the PPF's positive slope, so that complementarity becomes weaker.

For example, raising scientific output by one, given initial output rates of one, two, and three, respectively, increases the number of graduate students by only five, three, and two. A university that concentrates solely on graduate education and generates no cited bioscience papers can maintain 283 graduate students at the sample-mean input levels.

Our results suggest that it is beneficial for this student-oriented university to reallocate its resources toward science because both outputs then tend to rise. Modern bioscience research projects are not isolated from the entire research system, and project success often stimulates related research endeavors. Moreover, lab projects normally consist of several main steps and papers may be published as intermediate outputs after individual steps are completed. These papers serve as inputs to graduate education in the sense of providing research topics and tools with which graduate students can work.

6.1.3 Optimal Allocation of R&D Expenditures and Returns to R&D

6.1.3.1 Optimal Allocation of R&D Expenditures

As discussed in section 4.7.2, the coefficients of *AgPD* and *BioPD* in the university model reflect movements along the corresponding isocost curves. The statistical significance of these coefficients indicates inoptimality in the allocation between post-docs and non-post-doc inputs in the university's agriculture and biology programs. In both paper and graduate student equation estimates, these coefficients are significant and positive; specifically, they are 0.137 and 0.013 in the science equation and 7.677 and 0.500 in the graduate student equation, respectively. The implication is that neither the agricultural nor the biological R&D expenditures of the representative university are optimally allocated between post-docs and non-post-doc inputs, whether we are speaking of bioscience production or graduate education. In particular, too few post-docs are employed. Indeed, holding total agricultural R&D expenditure and all other variables fixed, a 1% increase in agricultural post-docs at the representative university will increase bioscience and graduate education outputs by

0.10% and 0.09%, respectively, at sample mean. Similarly, a 1% increase in post-docs in that university's biology program will bring 0.17% and 0.11% increases, respectively, in bioscience and graduate education outputs.

6.1.3.2 Returns to R&D

Tables 6.3 and 6.4 show the short-run and long-run effects of university R&D expenditures on bioscience and graduate education under alternative policy settings. The effects are computed, from the equations presented in section 4.7.2, in both marginal (slope) and elasticity form. The average post-doc salary is assumed to be $30,000/year in both the agricultural and biological sciences. All elasticities reported in the present chapter are evaluated at sample mean input and output levels.

I will first discuss the short-run returns to R&D in university bioscience production (top portion of table 6.3). Consider alternative scenarios in which the representative university invests an additional $10,000,000 unit of R&D expenditures:

(a1) solely in its agriculture program;

(a2) solely in its biology program; or

(a3) in a combination of these two.

The largest increase in the university's bioscience output is achieved when that money is allocated only to post-docs (horizontal movement in figure 4.2). In particular, allocating all new R&D money to post-docs increases cited bioscience output by 46.01, 4.51, and 18.45 more cited bioscience papers, given strategies (a1), (a2), and (a3), respectively. In contrast, if this additional R&D unit is allocated only to non-post-doc inputs (vertical movement in figure 4.2), the university's cited bioscience

Table 6.3. R&D Impacts on Cited Bioscience Production in U.S. Universities: Marginal Effects and Elasticities

	Short-Run					
	Ag Science		Bio Science		Ag & Bio Science	
Policy Setting	Elasticity	Marginal	Elasticity	Marginal	Elasticity	Marginal
Vertical	0.079	0.436	0.116	0.323	0.195	0.361
Horizontal	8.362	46.010	1.618	4.506	9.980	18.452
Average	0.173	0.952	0.290	0.808	0.463	0.856

	Long-Run					
	Ag Science		Bio Science		Ag & Bio Science	
Policy Setting	Elasticity	Marginal	Elasticity	Marginal	Elasticity	Marginal
Vertical	0.213	1.170	0.310	0.865	0.523	0.967
Horizontal	22.418	123.350	4.338	12.080	26.756	49.470
Average	0.464	2.551	0.778	2.167	1.242	2.296

Note 1. The post-doc salary is assumed to be $30,000/year in both agricultural and biological sciences.

2. All evaluations are at sample means.

Table 6.4. R&D Impacts on Graduate Education in U.S. Universities: Marginal Effects and Elasticities

	Short-Run					
	Ag Science		Bio Science		Ag & Bio Science	
Policy Setting	Elasticity	Marginal	Elasticity	Marginal	Elasticity	Marginal
Vertical	0.139	48.100	0.225	39.300	0.364	42.257
Horizontal	7.548	2606.993	1.178	205.880	8.726	1012.725
Average	0.223	77.039	0.336	58.645	0.559	64.826

Note 1. The post-doc salary is assumed to be $30,000/year in both agricultural and biological sciences.

2. All evaluations are at sample means.

output rises by only 0.44, 0.32, and 0.36 under the three respective scenarios. Alternatively, if the additional R&D unit is allocated between post-doc and non-post-doc inputs proportionately to their sample-mean share in R&D expenditures (average movement in figure 4.2), the university's bioscience output rises by 0.95, 0.81, and 0.86, respectively, namely the weighted sum of the corresponding vertical and horizontal effects.

The strikingly high marginal returns to investment in post-docs compared to that in non-post-doc inputs implies that agriculture and biology programs at U.S. universities heavily underinvest in post-doctoral researchers. This is especially serious in agricultural programs, where the marginal effect is as high as 46.01. Essentially, utilization of post-docs is in stage I of the university production function, where the marginal product of this input is extremely high. At sample means, post-docs account for only 1.1% of university agricultural R&D expenditures, and for only 11.6% in biological R&D. The important role post-docs play in a university's successful generation of bioscientific knowledge can be explained by their professional maturity gained in Ph.D. study, their strong incentive to publish, and their focus on research work. It seems that, especially in their agricultural programs, U.S. universities have not taken full advantage of post-docs' high research productivity.

Under each of the three allocation schemes between post-docs and non-post-doc inputs in table 6.3 (vertical, horizontal, or average), we find that investing the additional R&D dollar solely in agricultural research [scenario (a1)] always brings a higher return in its bioscience output than does investing in biology program research or in a combination of agriculture and biology. This is understandable because bioscience is measured here by the number of produced papers cited in agricultural biotechnology patents, and universities' biology programs probably produce more papers that are uncited in agbiotech patents than agriculture programs do.

In sum, if a university is given an additional $10,000,000 in R&D funding, the highest returns in cited bioscience papers would be attained if this money were invested in agricultural program post-docs, and the lowest returns if it were invested in biology program non-post-doc inputs.

Short-run elasticities reported in the top portion of table 6.3 reflect the returns to R&D in university bioscience production. Under all three alternative assumptions about the program in which the representative university invests its additional unit of R&D expenditure, increasing returns to scale are observed only when that money is allocated solely to post-docs. Specifically, a 1% increase in a university's R&D in agricultural science, biological science, or combination thereof induces, respectively, a 8.36%, 1.62%, or 9.98% increase in cited bioscience output, provided all the additional money is allocated to post-docs. If the additional money is allocated solely to non-post-doc inputs, strongly decreasing returns to scale (0.08, 0.12, and 0.2 in the three scenarios) predominate. If the additional R&D is distributed in proportion to the average allocation between post-docs and non-post-doc inputs, decreasing returns to scale, namely 0.17, 0.29, and 0.46, respectively, are also evident. Increasing returns to scale in R&D invested in post-docs reconfirms U.S. universities' incentive to reallocate expenditure toward post-doctoral fellows.

Although marginal effects in scenario (a1) are always greater than those in (a2) and (a3), elasticities do not necessarily follow the same pattern. Elasticities are marginal effects multiplied by the corresponding R&D-to-output ratios $AgRD / Science$, $BioRD / Science$, or $(AgRD + BioRD) / Science$. At sample mean, *BioRD* is approximately twice as large as *AgRD*, so that all three elasticities in scenario (a3) (namely those associated with the vertical, horizontal, and average movements in figure 4.2) are larger than those in (a1) and (a2). Hence, the largest short-run return to R&D in university bioscience is 18.45%, achieved when the additional R&D is invested entirely in post-docs in scenario (a3). In the same "average-allocation" scenario, however, we might more realistically assume the university spends the additional R&D dollar in proportion to the average allocation between post-docs and non-post-doc inputs. In that case, a 1% increase in R&D would bring about a 0.46% increase in university bioscience output.

Long-run measures of these marginal effects and elasticities, reported in the bottom portion of table (6.3), are the corresponding short-run measures multiplied by $1/(1-\lambda)$, where λ is estimated in table 6.1 to be 0.627. It is interesting that long-run average returns to R&D remain decreasing under scenarios (a1) and (a2), but become increasing under scenario (a3). That is, if the representative university's total R&D expenditures in agricultural and biology programs each increase by 1%, its bioscience output increases by 1.50% provided the university allocates the increased expenditures in such a way that the proportional share of post-docs and non-post-doc inputs in total spending remains constant.

Turning to the returns to R&D in university graduate education production, presented in the top portion of table 6.4, we observe a pattern similar to that in bioscience production. In all three scenarios (a1), (a2), and (a3), a $10,000,000 increase in university R&D induces a larger increase in graduate student numbers if the R&D increase is allocated only to post-docs than if it is allocated only to non-post-doc inputs or to a combination of the two. The respective marginal impacts are 2607, 206, and 1012 additional graduate students under these three regimes. The much greater marginal returns to post-docs than to non-post-doc inputs reported in table 6.4, this time in connection with graduate education, indicate again an under-expenditure on post-doctoral fellows in the agricultural and biology programs in U.S. universities. Many post-docs interact daily with graduate students' laboratory experiments and provide them with help and advice during the research process. Hence, post-docs serve not only as a research input but as an input to graduate education.

Comparing the three scenarios in table 6.4, we find that allocating investments solely to agricultural programs always brings a higher change in graduate student population than does allocating investment partly to biology programs, regardless of the allocation of this money between post-docs and non-post-doc inputs. This result implies that the marginal cost of graduate education is lower in agricultural than in biology programs. Agricultural programs in Land Grant universities, at least, traditionally have been more education-oriented than have science college programs. Because they presumably have a superior infrastructure for providing education, they would presumably enjoy lower marginal costs as well.

Returns to R&D scale in university graduate education are reflected in the elasticities reported in table 6.4. Again, in all three scenarios (a1), (a2), and (a3), that is regardless of the allocation of expenditure between agriculture and biology programs, increasing returns to scale occur only when the additional R&D is allocated solely to post-doctoral fellows. Increasing the representative university's post-doctoral expenditures by 1% in agricultural science, biological science, or both raises graduate output by 7.55%, 1.18%, or 8.73%, respectively. Assuming the sample-mean allocation between agricultural and biology programs is maintained, investing the additional R&D money in post-docs alone gives the largest percentage return, namely 8.73%. If in the same scenario the university spends the additional R&D on both post-docs and non-post-doc inputs, a 1% R&D increase would bring about only a 0.56% increase in graduate student population.

In summary, the impacts of R&D on cited bioscience and graduate education in U.S. universities, measured both in marginal (slope) and elasticity terms, differ significantly depending upon the allocation of R&D between agricultural and biology programs and between post-docs and other (faculty, equipment, and material) inputs. Policy choices therefore have much to do with the effectiveness of R&D investment.

6.1.4 Other Elasticities

Elasticities of other variables in the university model reflect their effects on university research and education productivities.

At sample mean input and output levels, the short-run elasticities of cited bioscience output with respect to a university's agricultural and biology program rankings are -0.02% (statistically nonsignificant) and -0.17%, respectively. That is, improvement in a university's quality ranking in those two fields reduces its production of bioscience papers cited by agbiotech patents, holding its R&D expenditure and other variables fixed. This result contradicts the generally held belief that leading universities are more efficient in performing research. Instead, it suggests that if the lower-ranked schools were provided the same amount of R&D money, they could produce more cited bioscience publications than do the higher-ranked ones. This can be because too much R&D expenditure is allocated to higher ranked universities, and decreasing returns to scale in R&D prevail there. However, the result can also be an illusion created by mismeasurement of university bioscience output. One possible source of mismeasurement is that leading research universities focus on higher quality research and bigger "hits", so that the raw, non-quality-weighted scientific paper counts used in the present study downward bias the bioscience output measures at those universities. Another possible source is that top universities focus more on basic research which, even though of long-term value, does not have direct applications to subsequently patented innovations. Thus, conclusions about the greater efficiency at lower ranked universities can be made only in terms of the total numbers of cited bioscience papers they produce.

Aggregate agricultural research expenditures in the private sector have positive and significant effects on universities' basic bioscience production. In particular, a 1% increase in private agricultural R&D leads to a 0.5% short-run increase in university's cited bioscience output. This spillover effect from private to public sector has been widely recognized in the economic literature. The short-run tradeoff between university basic bioscience and applied biotechnology is significant at the 90% level but small in magnitude. At a given R&D expenditure, a university's cited bioscience output drops by 0.009% when its agbiotech patent output rises by 1%. Hence, these two research outputs are substitutes rather than complements to one another.

Turning to a representative university's graduate education production, we observe that, everything else held constant, increasing the agricultural program ranking by one unit reduces graduate output by 0.58%, and increasing the biological program ranking by one reduces graduate output by 0.05%. This implies that higher-ranked schools have a higher R&D-to-graduate-student ratio, and thus a presumably higher average cost of graduate education. Again, the quality variation across universities' graduate educations is not captured in our graduate output measure, namely the total number of Masters and Ph.D. students in agricultural and biological sciences). Although highly ranked programs attract students, their admission standards are also unusually high and they restrict their graduate student populations to maintain quality. Moreover, highly ranked schools tend to focus on Ph.D. education, which is more research-oriented than is Masters education and thus more demanding in R&D funding.

A representative university's graduate output responds negatively but nonsignificantly to a marginal change in aggregate private-sector agricultural R&D. A 1% increase in private-sector R&D brings about a 0.04% decrease in graduate student numbers, likely because some potential graduate program candidates are attracted by positions in the private sector created by these increased R&D expenditures. A shift in a university's agenda toward applied biotech research significantly reduces its graduate student numbers. Specifically, a 1% increase in the number of agbiotech patents produced at a university leads to a 0.005% decrease in its graduate student numbers in agricultural and biological sciences.

6.2 Results of Biotechnology Firm Model Estimation

6.2.1 Parameter Estimates

Our biotechnology firm model consists of the two equations (4.10) and (4.11), in which a firm's patenting rates in, respectively, agricultural biotechnology and other fields are functions of its R&D expenditures and other inputs. The model is estimated using the unbalanced, pooled time-series cross-sectional data on 85 biotech firms from 1985 to 2000, as discussed in chapter 5. The unbalanced nature of this sample is caused by two factors. First, many of the firms existed under one name during only part of our study period, either because they did not exist during some of those years or because they had been merged into or acquired by other firms. Second, the R&D expenditures of some of the firms that did exist during particular years from 1985 to 2000 were not reported in the Census Bureau's industry R&D database. After the observations with such missing values were excluded, 730 observations were left for estimation purposes.

As discussed in section 4.2.2, a firm's R&D intensity, measured by the ratio of its total R&D expenditures to its total sales (*RDint*) or by the ratio of its total scientist and engineer numbers to its total employee numbers (*SEint*), was initially included as characteristics variables in our model. These two variables were, however, nonsignificant and did not improve the goodness-of-fit of either the agbiotech or non-agbiotech equation. The implication is that a firm's R&D intensity does not have any significant effect on its patent production, holding constant its total R&D expenditures and scientist and engineer force. The results reported below do not include these two variables. A firm's size can be measured by its total sales volume (*Sales*) or total employee number (*Employee*). Replacing either of these two variables by the other produces almost identical parameter estimates for the remaining variables in both equations. I will present only the results that include *Sales*.

The commercial firm agbiotech and non-agbiotech patent equations were, as in the university model, estimated in a number of alternative ways. First, they were estimated separately with (a) OLS, (b) a fixed effects model, and (c) a model in which a heteroskedasticity correction is made to the OLS estimates. Then, the equations were estimated jointly with SUR. Parameter estimates, their corresponding t-statistics, and the goodness-of-fit of the two equations using each indicated estimation technique, are presented in tables 6.5 and 6.6.

In the single-equation estimates, the R^2 of the agbiotech and non-agbiotech equations range from 0.57 to 0.77 and from 0.24 to 0.59, respectively. The system weighted R^2 in the SUR estimate is 0.47. Although the R^2s in the non-agbiotech equation are not as high as those in the agbiotech equation, most of its parameters are

Table 6.5. Agricultural Biotchnology Patent Production in Commercial Firms: Parameter Estimates

	OLS		OLS Hetero corrected		Fixed Effects		SUR	
Variable	Estimate	t	Estimate	t	Estimate	t	Estimate	t
Intercept	-6.200	-4.16**	-0.620	-2.79**	-5.660	-5.07**	-6.450	-4.32**
NonAgPt	-4.20E-03	-2.35**	1.20E-04	0.36	-1.60E-03	-1.07	-2.80E-03	-1.57
$NonAgPt^{0.5}$	0.137	3.74**	0.002	0.44	0.049	1.7*	0.135	3.72**
FirmRD	2.90E-03	0.73	1.02E-04	0.11	6.25E-04	0.19	1.83E-03	0.46
SE	-8.40E-05	-1.38	-7.77E-07	-0.05	-2.10E-05	-0.42	-1.10E-04	-1.76*
SciInput	0.028	27.88**	0.059	33.69**	0.033	28.41**	0.028	27.7**
Sales	-4.02E-05	-0.43	-4.24E-06	-0.34	4.25E-05	0.61	-4.08E-05	-0.43
t	0.065	4.11**	0.007	2.83**	0.059	5.01**	0.067	4.26**
R^2	0.57		0.63		0.77		0.47	

Note 1. ** and * indicate parameters significant at 95% and 90% confidence levels, respectively.

2. *FirmRD*, *Sales*, and *SE* are lagged three years; all other independent variables are current.

Table 6.6. Non-Agricultural-Biotechnology Patent Production in Commercial Firms: Parameter Estimates

	OLS		OLS Hetero corrected		Fixed Effects		SUR	
Variable	Estimate	t	Estimate	t	Estimate	t	Estimate	t
Intercept	250.280	3.13**	214.640	3.75**	49.160	0.77	272.800	3.41**
AgPt	-4.530	-1.52	1.810	0.7	3.470	1.3	-2.590	-0.87
$AgPt^{0.5}$	33.080	3.21**	4.630	0.62	-5.650	-0.63	32.440	3.15**
FirmRD	0.760	3.73**	2.000	6.37**	-0.370	-2.1**	0.740	3.64**
SE	0.018	5.62**	0.008	1.7*	0.029	11.35**	0.018	5.7**
Sales	5.99E-04	0.12	-1.32E-03	-0.38	-2.91E-03	-0.75	8.97E-04	0.18
t	-2.410	-2.81**	-2.120	-3.53**	-0.210	-0.3	-2.670	-3.11**
R^2	0.29		0.24		0.59		0.47	

Note 1. ** and * indicate parameters significant at 95% and 90% confidence levels, respectively.

2. *FirmRD*, *Sales*, and *SE* are lagged three years; all other independent variables are current.

estimated with high precision. Given the constraint in data availability, our model reflects at least part of the firms' patent production technology.

In both the agbiotech and non-agbiotech equations, the OLS and SUR estimators produce coefficient estimates that are rather similar to one another, and SUR does not improve estimation efficiency. That is, as found in the university model, no strong contemporaneous correlation appears to exist between the error terms of the two equations, and SUR does not clearly provide any advantage over OLS. Although the fixed effects model in tables 6.5 and 6.6 cannot be rejected in a joint F-test, most of the corresponding firm-specific effects (not shown in the table) are nonsignificant. The fixed effects also draw explanatory power away from the specified variables and induce the slope of *FirmRD* to become negative, violating monotonicity in this input. Hence, OLS is selected in favor of SUR and of the fixed effects model in the present study.

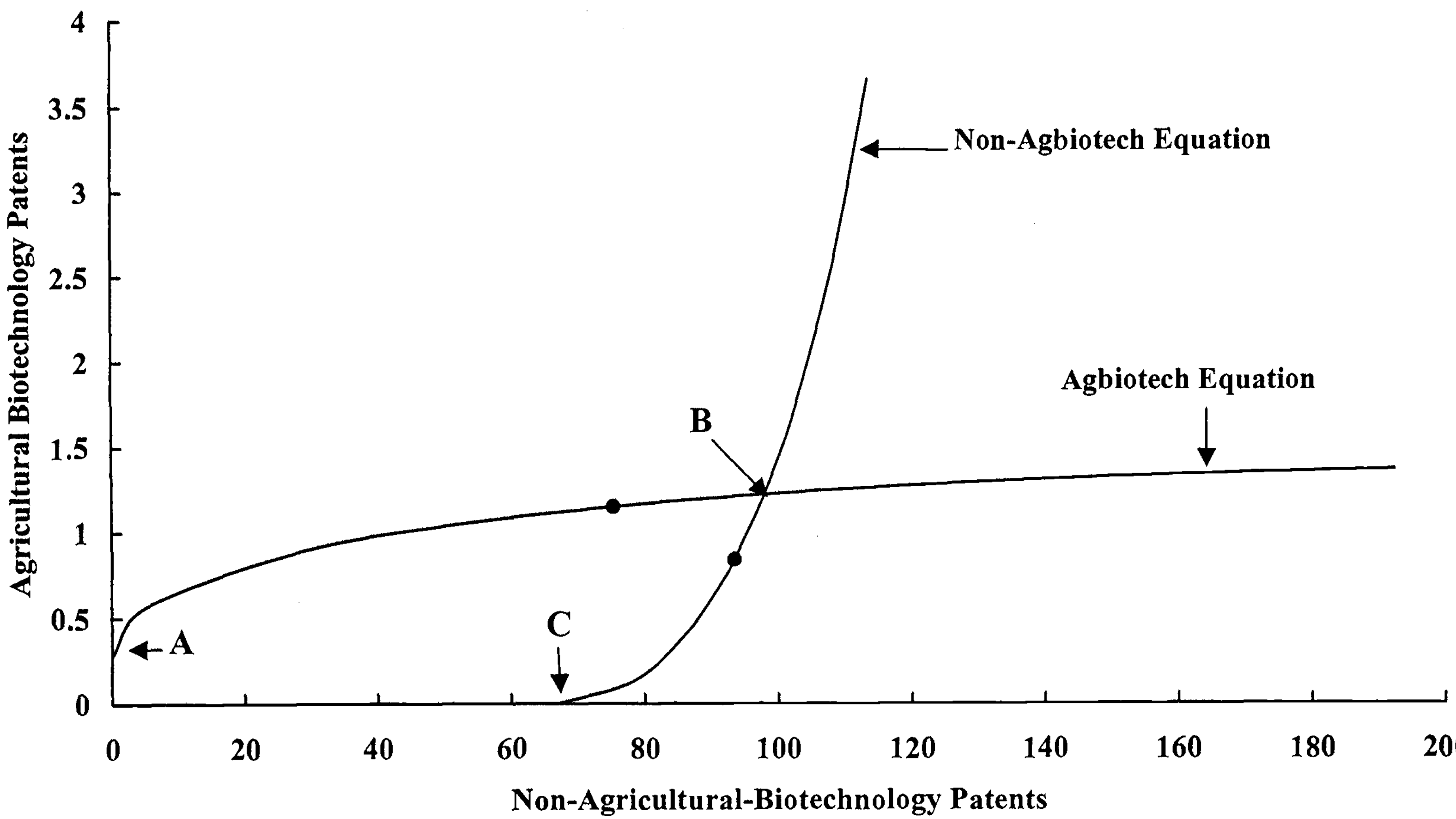
Heteorskedasticity is detected in the OLS estimation of both the agbiotech and non-agbiotech equations. Correcting for it changes parameter estimates but reduces their statistical significance. As discussed above, OLS estimates are, in the presence of heteorskedasticity, consistent although not efficient in large samples. Overall, statistical significance in the OLS estimates of the biotech firm model are high. In the agbiotech equation, all variables except *FirmRD*, *Sales*, and *SE* are significant at the 10% level. In the non-agbiotech equation, all variables except *AgPatent* and *Sales* are significant at the 10% level. In the following, I discuss the biotech firm model results based on the OLS estimation.

6.2.2 Production Possibility Frontier

The two biotech firm production possibility frontiers, based on the agbiotech and non-agbiotech patent equations respectively, are traced in figure 6.2. The explanatory output variables, namely, the number of non-agbiotech patents in the agbiotech equation and the number of agbiotech patents in the non-agbiotech equation, vary from zero to one standard deviation above their respective sample means, and all inputs are held fixed at sample means. Because the sample is right-skewed in these two outputs, especially in the number of agbiotech patents, the bulk of the university sample lies at the lower portion of the two PPF sections. For example, the sample median and mean of the quantity of agbiotech patents are 0 and 0.84, respectively, and the sample standard deviation is 2.82. The sample median and mean of the quantity of non-agbiotech patents are 30 and 75.34, respectively, and the sample standard deviation si 117.27. Hence, similar to the university model, our two biotech firm regressions are more accurately evaluated at lower output levels.

Analogous to the construction of the grand PPF in the university model, the joint set of the above two PPF sections, based on the two biotech firm equations, approximately represent the firms' "grand-technology" in producing agbiotech and non-agbiotech patents. In figure 6.2, the PPF portions from A to B and from C to B represent output combinations that a respectively agbiotech- and non-agbiotech-oriented firm can produce with sample-mean input levels. For the same reason as in the university model, a kink occurs at the intersection of the two PPF sections, and this

Figure 6.2. Agricultural Biotechnology and Non-Agricultural-Biotechnology Patent Output in Commercial Firms: the Production Possibility Frontier



portion of the grand PPF likely is not very accurate. But at points close to A and C, where most of the data points lie, the figure 6.2 representation should be quite reliable.

Firms represented between A and B in figure 6.2 are those specializing in agricultural biotechnology. Particularly near point A, the ratio of agbiotech to non-agbiotech patent production is high. This neighborhood of the PPF is positively sloped, so that non-agbiotech patents are complementary to agbiotech patents and thus reallocating resources from the first to the second increases both outputs. The complementarity initially is strong. Increasing non-agbiotech output from zero to one leads to an increase of 0.13 in agbiotech output. As reallocation proceeds, the positively sloped PPF becomes flatter and the complementarity becomes weaker. For example, when the non-agbiotech patent output rises from 40 to 41, agbiotech patent output rises by only 0.007.

A similar analysis applies to firms specializing in pharmaceutical and other technologies. These firms are represented by the PPF section between C and B in figure 6.2, where the ratio of agbiotech to non-agbiotech patent production is low. Between C and B, both non-agbiotech and agbiotech patent production rises as resources are reallocated form the former to the latter because the latter is an input to the former. Increasing agbiotech patent quantity from zero to 0.1 raises non-agbiotech patent production by 10. Further input reallocation toward agricultural biotechnology weakens the complementarity between the two outputs. For example, raising agbiotech output from 0.5 to 0.6 increases non-agbiotech patent output by only 2, down from the 10 patents near point C.

In summary, agricultural and non-agricultural biotechnology are complements to one another at both agbiotech and principally-non-agbiotech research firms. Research and development in those two fields are often joint: activity in one may enhance productivity in the other by providing it with research techniques, tools, and materials, sometimes called spillovers. This is especially so because agricultural and pharmaceutical biotechnologies are based on the same recombinant DNA principles and share many research procedures. A new promoter developed in a company's pharmaceutical research may be applicable to the genetic modification of a corn variety, while contents extracted from a GM crop or livestock program might prove useful in a human drug program.

6.2.3 Optimal Allocation of R&D Expenditures and Returns to R&D

6.2.3.1 Optimal Allocation of R&D Expenditures

As discussed in section 4.7.2, the coefficient of *SE* in the biotechnology firm model can be used to test the optimality of the firm's allocation between scientists/engineers and non-salary inputs. In the agricultural biotechnology equation, coefficient γ_5 is only –8.40E5 and nonsignificantly different from zero. This might indicate the representative firm's R&D expenditures are optimally allocated between salary and non-salary inputs in agbiotech patent production. But the nonsignificance may also be explained by randomness in the innovation process itself, that is by risks in the biotechnology research process. Nevertheless, the negative sign of γ_5 may instead imply that too many scientists and engineers are employed to minimize R&D

costs at given output levels. In that case, augmenting this human capital input under a fixed budget constraint would reduce agricultural biotechnology patent output.

The estimate of coefficient φ_5 in the non-agbiotech equation is 0.018 and significantly different from zero, indicating a non-cost-minimizing allocation of the representative firm's R&D between scientists/engineers and non-salary research inputs in non-agbiotech patent production. Indeed, holding total R&D expenditures and all other variables fixed, a 1% increase in the number of scientists and engineers (thus reducing non-salary inputs) increases the firm's non-agbiotech patent output by 0.31%. Commercial firms appear on average to underspend on human capital.

6.2.3.2 Returns to R&D

Impacts on the firm's agbiotech and non-agbiotech patent production of R&D expenditures under alternative policy settings are reported in marginal and elasticity form in table 6.7. The average corporate scientist/engineer salary is assumed here to be $60,000/year and all elasticities are evaluated at sample means. I assess alternative assumptions about how the firm allocates an additional unit of R&D expenditure between its two research input types. The alternative allocations are:

(b1) entirely on non-salary inputs;

(b2) entirely on scientists and engineers; or

(b3) on a combination of these two.

Table 6.7. R&D Impacts on Patent Production in U.S. Commercial Firms: Marginal Effects and Elasticities

Policy Setting	Agricultural Biotechnology Patent		Non-Agricultural Biotechnology Patent		Total Patent	
	Elasticity	Marginal	Elasticity	Marginal	Elasticity	Marginal
Vertical	0.085	0.003	0.248	0.760	0.246	0.763
Horizontal	-0.326	-0.011	1.212	3.710	1.195	3.699
Average	-0.046	-0.002	0.555	1.700	0.549	1.698

Note 1. The average corporate scientist and engineer salary is assumed to be $60,000/year.

2. All evaluations are at sample means.

The above three scenarios correspond to the policy settings referred to as horizontal, vertical, and average in table 6.7.

As estimates of the coefficients of the firm's R&D expenditures and scientist-and-engineer numbers, γ_4 and γ_5 respectively, are not significant in the agbiotech equation, marginal effects and elasticities based on them should be interpreted with caution. Increasing the representative firm's R&D expenditures by $10,000,000 raises its agricultural biotechnology patent output by 0.003 per year under policy (b1) and decreases this patent output by 0.011 and 0.002 under policies (b2) and (b3). These marginal effects are very small compared to the corresponding effects on non-agbiotech output, consistent with the insignificance of the parameter estimates.

Returns to R&D in the firm's non-agricultural-biotechnology patent production are positive and substantial. A $10,000,000 increase in R&D expenditures raises non-agbiotech patent output by 0.76, 3.71, or 1.7 when the additional R&D money is allocated respectively to non-salary inputs, to scientists and engineers, or to a combination of them. Thus, if the firm is to spend $10,000,000 more on R&D, it is optimal in terms of non-agbiotech output to spend it entirely on scientists and engineers, namely on scenario (b2). Elasticities are the corresponding marginal effects multiplied by the input-output ratio $FirmRD / NonAgPt$ evaluated at sample means. As table 6.7 shows, increasing returns to R&D scale prevails in scenario (b2), but decreasing returns predominate in scenarios (b1) and (b3).

Summing the marginal R&D effects on agbiotech and non-agbiotech patent output gives the marginal R&D effect on total patent output. As mentioned above, marginal effects on agbiotech output are very small compared to those on non-agbiotech output, so that total marginal effects are almost identical to those on non-agbiotech output. Multiplying the summed marginal effects by ratio $FirmRD \,/\, (AgPt + NonAgPt)$ gives the corresponding elasticities. As agbiotech patent output accounts, at sample means, for only 1.1% of total patent output, elasticities of total output with respect to R&D expenditure are also very close to those of non-agbiotech patent output.

6.2.4 Other Elasticities

Other factors affecting biotech firms' agbiotech and non-agbiotech patent production are also included in our model. While the effects of a firm's sales volume on its production of the two patent types are very nonsignificant, the contribution of scientific inputs to agbiotech inventions, and the change in a firm's propensity to patent both types of innovations, are significant.

The increase in a firm's agbiotech patent output brought about by a 1% increase in cited scientific papers is 0.55%. This result is consistent with other economists' findings about the positive spillovers from university research to the private sector. Indeed, especially in the increasingly science-based biotechnology, universities' unique capacity in basic research is essential in facilitating applied and patentable innovations.

Holding all else constant, a representative firm's propensity to patent agbiotech inventions increased by an average of 8% each year during the study period. In contrast, it's propensity to patent non-agbiotech inventions decreased by 3.2% each year during the same period. During the last two decades, many conventional chemical manufacturers have shifted their research and production emphasis to biotechnology, where technological opportunities have not been as fully exploited and where the urge to patent has become part of the industry culture.

6.3 Complementarity Between Basic Bioscience and Applied Biotechnology

Total complementarity between basic bioscience and applied biotechnology can be computed from equation (4.55), using alternative marginal impact measures of firm R&D on patent production and university R&D on bioscience production. These alternative measures correspond to the alternative policy settings we have been discussing. Substituting $\alpha_9 = 0.0066$ and $\gamma_3 = 0.028$ into (4.55) gives

$$\frac{d\ Pt}{d\ UnivRD} = -\frac{\partial\ Pt}{\partial\ FirmRD} + 0.028\left(\frac{\partial\ Science}{\partial\ UnivRD} - 0.0066\right) \tag{6.1}$$

where *UnivRD* refers alternately to the university's agricultural-program R&D only, its biology-program R&D only, or a combination of the two; and *Pt* refers alternately to the representative firm's agricultural biotechnology patent output or total patent output.

Table 6.8 shows the complementarity measures computed using alternative combinations of the nine short-run marginal effects on university bioscience reported in table 6.3 and the six marginal effects on commercial firm patent production reported in table 6.7. Corresponding measures using university long-run marginal effects are presented in table 6.9.

If we concentrate on the firm's agricultural biotechnology patent production only, the complementarity values in both tables are all positive, indicating the university's basic bioscience research and the firm's applied agricultural biotechnology research are always complements to one another. This is because the research money reallocated from the representative firm to the representative university has a significant positive effect on the university's bioscience output, whereas the reduction in the firm's R&D has no significant impact on its agbiotech patent production.

If we now expand our attention in tables 6.8 and 6.9 to total patent output, including that in pharmaceutical and other non-agricultural biotechnology, the incidence of complementarity between science and technology falls. In the short run (table 6.8), 15 of the 27 values shown are negative, indicating substitutability between basic science and applied technology. In the long run (table 6.9), however, 18 of the 27 values are positive, suggesting that complementarity is quite high even in non-agricultural research.

Table 6.8. Total Complementarity Between Basic and Applied Research, Assuming Short-Run Effects of Basic Research

		Agbiotech Patent			Total Patent		
		Vertical	Horizontal	Average	Vertical	Horizontal	Average
Ag Science	Vertical	0.009	0.023	0.014	-0.274	-3.210	-1.210
	Horizontal	1.285	1.299	1.290	50.768	47.832	49.833
	Average	0.017	0.038	0.028	0.303	-2.633	-0.632
Bio Science	Vertical	0.006	0.020	0.010	-0.402	-3.338	-1.337
	Horizontal	0.123	0.137	0.128	4.283	1.347	3.348
	Average	0.020	0.034	0.024	0.142	-2.794	-0.793
Ag & Bio Science	Vertical	0.007	0.021	0.011	-0.359	-3.295	-1.294
	Horizontal	0.514	0.528	0.518	19.904	16.968	18.968
	Average	0.021	0.035	0.025	0.196	-2.740	-0.739

Table 6.9. Total Complementarity Between Basic and Applied Research, Assuming Long-Run Effects of Basic Research

		Agbiotech Patent			Total Patent		
		Vertical	Horizontal	Average	Vertical	Horizontal	Average
Ag Science	Vertical	0.030	0.044	0.034	0.547	-2.389	-0.388
	Horizontal	3.451	3.465	3.455	137.389	134.453	136.454
	Average	0.042	0.082	0.073	2.095	-0.841	1.159
Bio Science	Vertical	0.021	0.035	0.026	0.205	-2.731	-0.730
	Horizontal	0.335	0.349	0.340	12.766	9.830	11.831
	Average	0.058	0.072	0.062	1.664	-1.272	0.729
Ag & Bio Science	Vertical	0.024	0.038	0.028	0.320	-2.616	-0.615
	Horizontal	1.382	1.396	1.387	54.643	51.707	53.708
	Average	0.061	0.075	0.066	1.809	-1.127	0.873

Chapter 7: Conclusions

Since the first wide-scale planting of GM crops in 1996, agricultural biotechnology has been adopted at a rapid rate in the United States and is profoundly reshaping the agricultural research sector. Broadly speaking, universities continue to focus on basic bioscience and commercial firms on applied biotechnology. Yet through the patenting process, both universities and firms find themselves in an increasingly privatized setting. More importantly, the increasingly science-based nature of biotechnology is providing an incentive for a closer working relationship between basic and applied research and between university and firm, and collaborations between the two have risen substantially. The collaboration ranges from research funding and scientist training to information and gene sequence exchange, material and equipment transfer, intellectual property licensing, and joint research endeavors. In the present dissertation, I have developed and analyzed a unique data set linking university bioscience with commercial agricultural biotechnology. My goal has been to characterize the impacts which each have on the other; more particularly, to specify and estimate the knowledge production relationships whereby science and technology interact to generate useful innovations.

The study's university model provides insights into the bioscience research and agricultural and biological graduate programs in U.S. universities. Results suggest that bioscience and graduate education are complements for one another at student-oriented universities, and slight substitutes for one another at research-oriented universities. Any shift in a university's emphasis toward applied, patentable agricultural biotechnology research draws scarce resources away from basic bioscience research and graduate education, reducing these latter two services. Surprisingly, highly ranked universities are less efficient in producing cited bioscience and graduate students than are their lower-ranked counterparts. This may, perhaps, be related to the non-quality-weighted manner in which science and education are deliberately measured in this study. However, the result is not unreasonable: if highly ranked universities produce more cited bioscience and graduate students than do others, it is not necessarily because they are more efficient in using their resources but simply because they acquire more of these resources. Aggregate private-sector agricultural R&D has a significant positive spillover effect on the representative university's basic bioscience research but a nonsignificant effect on its agricultural and biological graduate education programs.

R&D expenditures in both agriculture and biology programs at U.S. universities appear to be inoptimally allocated between post-doctoral fellows and non-post-doctoral inputs. In particular, universities heavily under-invest in post-doctoral fellows, and greater returns to R&D would be attained if additional R&D funding were invested in post-docs. Allocating another R&D dollar solely to the university's agricultural programs brings a higher increase in cited bioscience papers and in graduate student enrollments than does allocating that dollar solely to biology programs or to a combination of agriculture and biology.

Results of the commercial-firm biotechnology model suggest that, both in agbiotech-only firms and in those with substantial non-agbiotech (primarily pharmaceutical) research, agricultural and non-agricultural biotechnology are complements to one another. Complementarities of such sort often are referred to as within-firm spillovers. The propensity of biotech firms in our sample to patent agricultural biotechnology inventions has significantly increased, while their propensity to patent in pharmaceutical and other areas has dropped. By explicitly linking, through the citation trail, universities' basic research with biotechnology firms' applied research, I find that university bioscience research plays an important and essential role in facilitating patentable agricultural biotechnology innovations.

Research and development in firms' agricultural biotechnology programs is, on average, nearly optimally allocated between scientists/engineers and non-salary research inputs. However, firms appear to under-spend on scientists/engineers in their pharmaceutical and other non-agricultural activities. Boosting a representative biotech firm's R&D expenditures would bring only a small change in its agbiotech patent output but a large change in its non-agbiotech patent output, especially if the increased funds were spent entirely on scientists and engineers. Because agricultural biotechnology patents account for only a small proportion of a representative biotech firm's total patent portfolio, conclusions with regard to the firm's non-agbiotech patent production apply equally well to the analysis of the firm's aggregate total patent production.

More broadly, we are concerned with the impacts of reallocating R&D expenditures across society at large, that is between universities and firms and between basic bioscience and applied biotechnology. Such impacts depend on how universities and biotech firms allocate any additional R&D funds obtained. The impacts also differ between agricultural and non-agricultural biotechnology research. In agricultural biotechnology alone, basic bioscience and applied biotechnology appear always to be complementary with one another. In pharmaceutical and other non-agricultural research, bioscience and applied technology are either complements or substitutes, depending upon how research monies are allocated between scientist and non-scientist inputs.

The predominantly complementary relationships identified in this research between science and graduate education, agricultural and pharmaceutical biotechnology, and science and technology in general, have implications for the design of research institutions and the walls between university and firm. For example, policies that encourage education-oriented universities to engage in more basic bioscience research would promote these universities' educational programs as well as contributing to technical innovation. Similarly, we have strong evidence that, on average, agricultural biotechnology programs are pursued most effectively in conjunction with pharmaceutical and other non-agricultural biotech programs, no doubt because of these programs' common link with gene transfer and cell regeneration methods.

Finally and most importantly, evidence of the complementarity between basic bioscience and applied biotechnology suggests that productivity in each of these two endeavors would be enhanced if communication between the two were strengthened. Likely sources of the complementarity are the science laboratory's growing demand for patented materials, processes, and equipment, and the technologist's continual need for a better understanding of molecular function. Institutions fostering improved communication include industrial R&D parks in the vicinity of research universities, departments that house scientists and technologists under a common roof, and short-term assignment of scientist-technologist teams to specific projects. Tax policies may also be designed to encourage private firms to collaborate with universities, and government grant programs may offer priority to joint university-firm proposals.

Bibliography

AUTM Licensing Survey, various years.

Acs, Z., and D. Audretsch (1988), "Innovation in Large and Small Firms: An Empirical Analysis." *American Economic Review*, vol.78: 678-90.

Acs, Z., D. Audretsch, and M. Feldman (1992), "Real Effects of Academic Research: Comment." *American Economic Review*, vol.82: 363-67.

Acs, Z., D. Audretsch, and M. Feldman (1994), "R&D Spillovers and Recipient Firm Size." *Review of Economics and Statistics*, vol.76: 336-40.

Adams, J. (1990), "Fundamental Stocks of Knowledge and Productivity Growth." *Journal of Political Economy*, vol.98: 673-702.

Adams, J., E. Chiang, and J. Jensen (2000), "The Influence of Federal Laboratory R&D On Industrial Research." NBER Working Paper No. 7612.

Adams, J., E. Chiang, and K. Starkey (2000), "Industry-University Cooperative Research Centers." NBER Working Paper No. 7843.

Adams, J., and Z. Griliches (1996), "Research Productivity in a System of Universities." NBER Working Paper No. 5833.

Aghion, P., and P. Howitt (1992), "A Model of Growth Through Creative Destruction." *Econometrica*, vol.60: 323 – 351.

Alston, J.M., G.W. Norton, and P.G. Pardey (1995), *Science Under Scarcity: Principles and Practice for Agricultural Research Priority Setting and Evaluation*. Cornell University Press, Ithaca, NY.

Arora, A., and A. Gambardella (1994a), "Evaluating Technological Information and Utilizing it: Scientific Knowledge, Technological Capability, and External Linkages in Biotechnology." *Journal of Economic Behavior and Organization*, vol.24: 91-114.

Arora, A., and A. Gambardella (1994b), "The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labor." *Research Policy*, vol.23: 523-32.

Audretsch, D., and M. Feldman (1996), "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review*, vol.86: 630-40.

Audretsch, D., and P. Stephan (1996), "Company-Scientist Locational Links: The Case of Biotechnology." *American Economic Review*, vol.86: 641-52.

Barham, B., J. Foltz, and K. Kim (2001), "Trends in University Ag-Biotech Patent Production." Research Report No. 58, Department of Agricultural and Resource Economics, University of Connecticut.

Basberg, B. (1982), "Technological Change in the Norwegian Whaling Industry: A Case Study of the Use of Patents Statistics as a Technology Indicator." *Research Policy*, vol: 11(3): 163-71.

Becker, G. (1962), "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, vol.70: 39-49.

Blundell, R., R. Griffith, and J. Reenen (1999), "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms." *Review of Economic Studies*, vol.66: 529-54.

Bredahl, M., and W. Peterson (1976), "The Productivity and Allocation of Research: U.S. Agricultural Experiment Stations." *American Journal of Agricultural Economics*, vol.58: 684-92.

Brennan, M., C. Pray, and A. Courtmanche (1999), "Impact of Industry Concentration on Innovation in the U.S. Plant Biotech Industry." Paper presented at the NE-165 Conference on Transitions in AgBiotech, Washington, D.C., June 1999.

Caswell, M., K. Fuglie, and C. Klotz (1994), *Agricultural Biotechnology: An Economic Perspective*. AER – 687, Economic Research Service, USDA.

Cockburn, I., and Z. Griliches (1988), "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R&D and Patents." *American Economic Review*, vol. 78: 419-23.

Cockburn, I., R. Henderson, and S. Stern (1999), "Balancing Incentives: The Tension Between Basic and Applied Research." NBER Working Paper No. 6882.

Cohen, W., and D. Levinthal (1989), "Innovation and Learning: The Two Faces of R&D." *The Economic Journal*, vol.99: 569-96.

David, P., and B. Hall (2000), "Heart of Darkness: Modeling Public-Private Funding Interactions Inside the R&D Black Box." NBER Working Paper No. 7538.

David, P., D. Mowery, and W. Steinmueller (1992), "Analysing the Economic Payoffs from Basic Research." *Economics of Innovations and New Technology*, vol.2: 73-90.

Dasgupta, P., and P. David (1994), "Toward a New Economics of Science," *Research Policy*, vol.23: 487-521.

Dasgupta, P., and E. Maskin (1987), "The Simple Economics of Research Portfolios," *Economic Journal*, vol.97: 581-95.

David, P., D. Mowery, and W. Steinmueller (1992), "Analyzing the Economic Payoffs from Basic Research." *Economics of Innovation and New Technology*, vol. 2: 73-90.

Eaton, J., and S. Kortum (1996), "Trade in Ideas: Patenting & Productivity in the OECD." *Journal of International Economics*, vol. 40: 251-78.

Ehrenberg, R. "The Flow of New Doctorates." *Journal of Economic Literature*, vol.30: 830-75.

Evenson, R. (1989), "Spillover Benefits of Agricultural Research: Evidence from U.S. Experience." *American Journal of Agricultural Economics*, vol.71: 447-52.

Foltz, J., B. Barham, and K. Kim (2000), "Universities and Agricultural Biotechnology Patent Production." *Agribusiness*, vol.16(1): 82-95.

Foltz, J., K. Kim, and B. Barham (2001), "A Dynamic Count Data Analysis of University Ag-Biotech Patents." Research Report No. 56, Department of Agricultural and Resource Economics, University of Connecticut.

Fuglie, K., N. Ballenger, K. Day, C. Klotz, M. Ollinger, J. Reilly, U. Vasavada, and J. Yee (1996), *Agricultural Research and Development: Public and Private Investments Under Alternative Markets and Institutions*. AER – 735, Economic Research Service, USDA.

Graff, G., G. Rausser, and A. Small (2001), "Agricultural Biotechnology's Complementary Intellectual Assets." Department of Agricultural and Resource Economics, University of California, Berkeley.

Graff, G., A. Heiman, and D. Zilberman (2001), "University Research, Offices of Technology Transfer, and Commercial R&D." Department of Agricultural and Resource Economics, University of California, Berkeley.

Griliches, Z. (1958), "Research Costs and Social Returns: Hybrid Corn and Related Innovations." *Journal of Political Economy*, vol.66: 419-31.

Griliches, Z. (1964), "Research Expenditures, Education, and the Aggregate Agricultural Production Function." *American Economic Review*, vol.54: 961-74.

Griliches, Z. (1967), "Distributed Lags, A Survey." *Econometrica*, vol.35: 16-47.

Griliches, Z. (1979), "Issues in Assessing the Contribution of R&D to Productivity Growth." *The Bell Journal of Economics*, vol.10: 92-116.

Griliches, Z. (1981), "Market Value, R&D, and Patents." *Economics Letters*, vol.7: 183-87.

Griliches, Z. ed. (1984), *R&D, Patents, and Productivity*. Chicago: University of Chicago press.

Griliches, Z. (1986), "Productivity, R&D, and Basic Research at the Firm Level in the 1970's." *American Economic Review*, vol.76: 141-54.

Griliches, Z. (1988), "Productivity Puzzles and R&D: Another Nonexplanation." *Journal of Economic Perspectives*, vol.2: 9-21.

Griliches, Z. (1990), "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, vol.28: 1661-707.

Griliches, Z. (1991), "The Search for R&D Spillovers." *Scandinavian Journal of Economics*, vol.94: 29-47.

Griliches, Z. (1994), "Productivity, R&D, and the Data Constraint." *American Economic Review*, vol.84: 1-23.

Griliches, Z., B. Hall, and A. Pakes (1991), "R&D, Patents, and Market Value Revisited: Is There a Second (Technological Opportunity) Factor?" *Economics of Innovation and New Technology*, vol.1: 183-202.

Hall, B. (1993), "The Stock Market's Valuation of R&D Investment During the 1980's." *American Economic Review*, vol.83: 259-64.

Hall, B. (2002), "The Financing of Research and Development." NBER Working Paper No. 8773.

Hall, B., Z. Griliches, and J. Hausman (1986), "Patents and R&D: Is There a Lag?" *International Economic Review*, vol.27: 265-83.

Hall, B., A. Jaffe, and M. Trajtenberg (2000), "Market Value and Patent Citations: A First Look." NBER Working Paper No. 7741.

Hall, B., A. Jaffe, and M. Trajtenberg (2001), "The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools." NBER Working Paper No. 8498.

Hall, B., A. Link, and J. Scott (2000), "Universities as Research Partners." NBER Working Paper No. 7643.

Harhoff, D., F. Narin, F. Scherer, and K. Vopel (1999), "Citation Frequency and the Value of Patented Inventions." *Review of Economics and Statistics*, vol. 81: 511-15.

Hausman, J., B. Hall, and Z. Griliches (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, vol.52: 909-38.

Henderson, R., A. Jaffe, and M. Trajtenberg (1998), "University as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988." *Review of Economics and Statistics*, vol. 80: 119-27.

Hayenga, M. (1998), "Structural Change in the Biotech Seed and Chemical Industrial Complex." *AgBioForum*, vol.1: 43-55.

Huffman, W.E., and R.E. Evenson (1989), "Supply and Demand Functions for Multiproduct U.S. Cash Grain Farms: Biases Caused by Research and Other Policies." *American Journal of Agricultural Economics*, vol.71: 761-73.

Huffman, W.E., and R.E. Evenson (1992), "Contributions of Public and Private Science and Technology to U.S. Agricultural Productivity." *American Journal of Agricultural Economics*, vol.74: 751-56.

Huffman, W.E., and R.E. Evenson (1993), *Science for Agriculture: A Long-Term Perspective*. Ames: Iowa State University Press.

Huffman, W.E., and R.E. Just (1994), "Funding, Structure, and Management of Public Agricultural Research in the United States." *American Journal of Agricultural Economics*, vol.76: 744-59.

Jaffe, A. (1986), "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review*, vol.76: 984-99.

Jaffe, A. (1988), "Demand and Supply Influences in R&D Intensity and Productivity Growth." *Review of Economics and Statistics*, vol.70(3): 431-37.

Jaffe, A. (1989), "Real Effects of Academic Research." *American Economic Review*, vol.79: 984-99.

Jaffe, A. (1999), "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process." NBER Working Paper No. 7280.

Jaffe, A., M. Trajtenberg, and M.S. Fogarty (2000), "The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees." NBER Working Paper No. 7631.

Jaffe, A., M. Trajtenberg, and R. Henderson (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, vol.108: 577-98.

Kalaitzandonakes, N., and M. Hayenga (1999), "Structural Change in the Biotechnology and Seed Industrial Complex: Theory and Evidence." Paper presented at the NE-165 Conference on Transitions in AgBiotech, Washington, D.C., June 1999.

Koo, B., and B. Wright (1999), "Dynamic Implications of Alternative Intellectual Property Rights for Crop Genetic Resources." Paper presented at the NE-165 Conference on Transitions in AgBiotech, Washington, D.C., June 1999.

Lanjouw, J., and M. Schankerman (1999), "The Quality of Ideas: Measuring Innovation with Multiple Indicators." NBER Working Paper No. 7345.

Levin, S., and P. Stephan (1991), "Research Productivity Over the Life Cycle: Evidence for Academic Scientists." *American Economic Review*, vol.81: 114-32.

Leslie, L., and R. Oaxaca (1993), "Scientist and Engineer Supply and Demand," in *Higher Education: Handbook of Theory and Research*, vol. IX: 154-211. Agathonn Press, New York.

Lichtenberg, F., and D. Siegel (1991), "The Impact of R&D Investment on Productivity: New Evidence Using Linked R&D-LRD Data." *Economic Inquiry*, vol. 29(2): 203-29.

Lyu, S., F. White, and Y. Lu (1984), "Estimating Effects of Agricultural Research and Extension on Productivity: A Translog Production Function Approach." *Southern Journal of Agricultural Economics*, vol.16: 1-8.

Mansfield, E. (1980), "Basic Research and Productivity Increase in Manufacturing." *American Economic Review*, vol.70: 863-73.

Mansfield, E. (1991), "Academic Research and Industrial Innovation." *Research Policy*, vol:20: 1-12.

Mansfield, E. (1995), "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing." *Review of Economics and Statistics*, vol. 77: 55-65.

McCunn A., and W.E. Huffman (2000), "Convergence in U.S. Productivity Growth for Agriculture: Implications of Interstate Research Spillovers for Funding Agricultural Research." *American Journal of Agricultural Economics*, vol.82: 370-88.

Merton, R. (1957), "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociology Review*, vol.22: 635-59.

Moschini, G., and H. Lapan (1997), "Intellectual Property Rights and the Welfare Effects of Agricultural R&D." *American Journal of Agricultural Economics*, vol.79: 1229-42.

Narin, F., K.S. Hamilton, and D. Olivastro (1997), "The Increasing Linkage Between U.S. Technology and Public Science." *Research Policy*, vol. 26, n3: 317-30.

Nelson, R. (1959), "The Simple Economics of Basic Scientific Research." *Journal of Political Economy*, vol.67: 297-306.

Nelson, R. (1986), "Institutions Supporting Technical Advance in Industry." *American Economic Review*, vol.76: 186-89.

Pakes, A. (1985), "On Patents, R&D, and the Stock Market Rate of Return." *Journal of Political Economy*, vol.93: 390-409.

Oehmke, J., D. Weatherspoon, C. Wolf, A. Naseem, M. Maredia, and A. Hightower (2000), "Is Agricultural Research Still A Public Good?" *Agribusiness: An International Journal*, vol.16: 68-81.

Pakes, A., and Z. Griliches (1980), "Patents and R&D at the Firm Level: A First Look." *Economics Letters*, vol.5: 377-81.

Pendleton, L., and J. Shonkwiler (2001), "Valuing Bundled Attributes: A Latent Characteristics Approach." *Land Economics*, vol.77, n1: 118-29.

Pardey, P. (1989), "The Agricultural Knowledge Production Function: An Empirical Look." *Review of Economics and Statistics*, vol.71: 453-61.

Porter, M., and S. Stern (2000), "Measuring the 'Ideas' Production Function: Evidence From International Patent Output." NBER Working Paper No. 7891.

Rausser, G. (1999), "Private/Public Research: Knowledge Assets and Future Scenarios." Invited Fellows Address, American Agricultural Economics Association Annual Meetings, Nashville, TN.

Ruttan, V. (1982), *Agricultural Research Policy*. University of Minnesota Press, Minneapolis.

Shoemaker, R., ed. (2001), *Economic Issues in Agricultural Biotechnology*. Agr. Info. Bull. No. 762, Economic Research Service, USDA.

Stern, S. (1999), "Do Scientists Pay to Be Scientists?" NBER Working Paper No. 7410.

Terleckyj, N. (1974), *Effects of R&D on the Productivity Growth of Industries: An Exploratory Study*. National Planning Association, Washington, D.C.

Thursby, J., and M. Thursby (2000), "Who is Selling the Ivory Tower? Sources of Growth in University Licensing." NBER Working Paper No. 7718.

Trajtenberg, M. (1990), "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *Rand Journal of Economics*, vol.21: 172-87.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997), "University Versus Corporate Patents: A Window on the Basicness of Invention." *Economics of Innovation and New Technology*, vol.5: 19-50.

Walsh, V. (1984), "Invention and Innovation in the Chemical Industry: Demand Pull or Discovery Push?" *Research Policy*, vol. 13(4): 211-34.

White, F., and J. Havlicek (1979), "Rates of Return to Agricultural Research and Extension in the Southern Region." *Southern Journal of Agricultural Economics*, vol.11: 107-11.

White, F., and J. Havlicek (1982), "Optimal Expenditures for Agricultural Research and Extension: Implications of Underfunding." *American Journal of Agricultural Economics*, vol.64: 47-55.

Zilberman, D., C. Yarkin, and A. Heiman (1999), "Agricultural Biotechnology: Economic and International Implications." Department of Agricultural and Resource Economics, University of California, Berkeley.

Zilberman, D., C. Yarkin, and A. Heiman (2001), "Privatization and Innovation in Agricultural Biotechnology." Department of Agricultural and Resource Economics, University of California, Berkeley.

Zucker, L., and M. Darby (1995), "Virtuous Circles of Productivity: Star Bioscientists and the Institutional Transformation of Industry." NBER Working Paper No. 5342 (www.nber.org).

Zucker, L., M. Darby, and M. Brewer (1998), "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, vol. 88: 290 – 306.

Zucker, L., M. Darby, and J. Armstrong (2002), "Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology." NBER Working Paper No. 8499.

APPENDICES

Appendix A: Second Approach for Deriving the Estimable Knowledge Production Functions

Starting from the same representation of the technology of biotechnological change given in Section 3.1.2,

$$Y_a^f = F_a^f\ (K_a^f, L_a^f, Y_b^u, X^f) \tag{3.1}$$

$$Y_b^u = F_b^u\ (K_b^u, L_b^u, K_a^f, L_a^f, X^u) \tag{3.2}$$

Let us jointly maximize over (K_a^f, L_a^f) and (K_b^u, L_b^u), Y_a^f as the final output and Y_b^u as the intermediate input, while (E_a^f, E_b^u) are given exogenouly:

$$\begin{aligned}
&\underset{K_a^f, L_a^f, K_b^u, L_b^u}{Max} && F_a^f\ (K_a^f, L_a^f, Y_b^u, X^f) \\
&s.t. && (W_k)_a^f K_a^f + (W_l)_a^f L_a^f \leq E_a^f \\
& && Y_b^u = F_b^u\ (K_b^u, L_b^u, K_a^f, L_a^f, X^u) \\
& && (W_k)_b^u K_b^u + (W_l)_b^u L_b^u \leq E_b^u
\end{aligned} \tag{A1}$$

At the optimal allocation of (K_a^f, L_a^f) and (K_b^u, L_b^u) obtained from the first order conditions of (A1), the maximized Y_a^f and Y_b^u are achieved at the given level of (E_a^f, E_b^u):

$$Y_a^f = G_a^{f'} \left(\frac{(W_k)_a^f}{E_a^f}, \frac{(W_l)_a^f}{E_a^f}, Y_b^u, X^f \right) \tag{A2}$$

$$Y_b^u \quad = \quad G_b^{u'} \left(\frac{(W_k)_b^u}{E_b^u}, \frac{(W_l)_b^u}{E_b^u}, \frac{(W_k)_a^f}{E_a^f}, \frac{(W_l)_a^f}{E_a^f}, X^f \right) \tag{A3}$$

Assume $(W_k)_a^f = (W_k)_b^u$ and $(W_l)_a^f = (W_l)_b^u$. Holding these prices fixed will allow us to eliminate them and rewrite the above two equations as

$$Y_a^f \quad = \quad G_a^f \left(E_a^f, Y_b^u, X^f \right) \tag{A4}$$

$$Y_b^u \quad = \quad G_b^u \left(E_b^u, E_a^f, X^u \right) \tag{A5}$$

Now we can find the socially optimal allocation of (E_a^f, E_b^u) by maximizing Y_a^f as the final output, using Y_b^u as the intermediate input and (E_a^f, E_b^u) as the choice variables. Substituting (A5) into (A4), we have

$$\begin{aligned} &\underset{E_a^f, E_b^u}{Max} \quad G_a^f \left(E_a^f,\ G_b^u \left(E_b^u, E_a^f, X^u \right),\ X^f \right) \\ &\quad s.t. \quad E_a^f + E_b^u \leq E^0 \end{aligned} \tag{A6}$$

When society allocates (E_a^f, E_b^u) optimally, the maximum of Y_a^f can be expressed as

$$Y_a^f \quad = \quad H \left(E^0, X^f, X^u \right) \tag{A7}$$

Therefore, the estimable knowledge production functions (A4) and (A5) derived with this approach are the same as those in equations (3.7) and (3.8').

Appendix B: The Front Page of an Agricultural Biotechnology Patent

United States Patent	*5,484,956*
Lundquist , et al.	**January 16, 1996**

Fertile transgenic Zea mays plant comprising heterologous DNA encoding Bacillus thuringiensis endotoxin

Abstract

Fertile transgenic Zea mays (corn) plants which stably express heterologous DNA which is heritable are provided along with a process for producing said plants. The preferred process comprises the microprojectile bombardment of friable embryogenic callus from the plant to be transformed. The process may be applicable to other graminaceous cereal plants which have not proven stably transformable by other techniques.

Inventors: **Lundquist; Ronald C.** (Minnetonka, MN); **Walters; David A.** (Bloomington, MN)
Assignee: **DeKalb Genetics Corporation** (DeKalb, IL)
Appl. No.: **508045**
Filed: **April 11, 1990**

Current U.S. Class:	**800/302**; 536/23.71
Intern'l Class:	A01H 004/00; C12N 015/05
Field of Search:	800/200,205,230 435/172.1,172.3,240.4,240.5,240.45 536/23.71

References Cited **[Referenced By]**

U.S. Patent Documents

4370160	Jan., 1983	Ziemelis	504/323.
4559302	Dec., 1985	Ingolia	435/172.
4581847	Apr., 1986	Hibberd et al.	47/58.
4665030	May., 1987	Close	435/240.
4666844	May., 1987	Cheng	435/240.
4727028	Feb., 1988	Santerre et al.	435/240.
4761373	Aug., 1988	Anderson et al.	435/172.
4806483	Feb., 1989	Wang	435/240.
4940835	Jul., 1990	Shah et al.	800/205.
5015580	May., 1991	Christou et al.	435/172.
5049500	Sep., 1991	Arnizen et al.	435/172.

Foreign Patent Documents

331853	., 0000	EP.
126537	Apr., 1983	EP.

141373	May., 1985	EP.
154204	Sep., 1985	EP.
160390	Nov., 1985	EP.
0193259	Sep., 1986	EP
202668	Nov., 1986	EP.
204549	Dec., 1986	EP.
242236	Oct., 1987	EP.
242246	Oct., 1987	EP.
299552	Jan., 1988	EP.
262971	Apr., 1988	EP.
275069	Jul., 1988	EP.
270356	Aug., 1988	EP.
280400	Aug., 1988	EP.
289479	Nov., 1988	EP.
292435	Nov., 1988	EP.
301749	Feb., 1989	EP.
334539	Sep., 1989	EP.
348348	Dec., 1989	EP.
442174	Aug., 1991	EP.
3738874	Nov., 1988	DE.
61-134343	Jun., 1986	JP.
8801444	Jan., 1990	NL.
2159173	Nov., 1985	GB.
WO85/01856	May., 1985	WO.
WO85/02972	Jul., 1985	WO.
WO87/05629	Sep., 1987	WO.
WO89/04371	May., 1988	WO.
WO89/12102	Dec., 1989	WO.
WO90/10691	Sep., 1990	WO.

Other References

Agricell Report, news article entitled "`Bullets` Transform Plant Cells", p. 5 (Jul. 1987).

H. Ahokes, "Electrophoretic Transfection of Cereal Grains with Exogenous Nucleic Acid", Soc. Biochem. Biophys., Microbio, Fen., Biotieteen Paivat (Bioscience Days), Abstracts, Technical University of Helsinki, Espoo, p. 2 (1989).

C. Armstrong et al., "Genetic and Cytogenetic Variation in Plants Regenerated from Organogenic and Friable, Embryogenic Tissue Cultures of Maize", Biol. Abstracts, vol. 85, Abstract 117662 (1988).

R. Barker et al., "Nucleotide Sequence of the T-DNA Region from the Agrobacterium tumefaciens octopine Ti plasmid pTi15955", Plant Mol. Biol.,

2:335-350 (1983).
M. Bevan et al., "A Chimearic Antibiotic Resistance Gene as a Selectable Marker for Plant Cell transformation", Nature, 304:184-187 (1983).
Cao et al., "Transformation of Rice and Maize Using the Biolistic Process", Plant Gene Transfer, 129:21-33 (1990).
V. Chandler et al., "Two Regulatory Genes of the Maize Anthocyanin Pathway Are Homologous: Isolation of B Utilizing R Genomic Sequences", The Plant Cell, 1:1175-1183 (1989).
Creissen et al., "Agrobacterium-and microprojectile-Mediated Viral DNA Delivery into Barley Microspore-Derived Cultures", Plant Cell Reports, 8:680-683 (Apr. 1990).
Christou et al., "Cotransformation Frequencies of Foreign Genes in Soybean Cell Cultures", Theor. Appl. Genet., 79:337-341 (1990).
A. Crossway et al., "Integration of Foreign DNA following Microinjection of Tobacco Mesophyll Protoplasts", Mol. Gen. Genet., 202: 179-185 (1986).
M. De Block et al., "Engineering Herbicide Resistance in Plants by Expression of a Detoxifying Enzyme", EMBO J., 6:2513-2518 (1987).
W. De Greef et al., "Evaluation of Herbicide Resistance in Transgenic Crops Under Field Conditions", Bio/Technology, 7:61-64 (1989).
R. Dekeyser et al., "Evaluation of Selectable Markers for Rice Transformation", Plant Physiol., 90:217-223 (1989).
DeWald et al. "Plant Regeneration from Inbred Maize Suspensions", Abstract Al-36, 7th International Congress on Plant Tissue and Cell Culture, Amsterdam, Jun. 24-29 (1990).
P. Evans, "Somaclonal Variation-Genetic Basis and Breeding Applications", Trends Genet., 5:46-50 (1989).
P. Fransz, Cytodifferentiation During Callus Initiation and Somatic Embryogenesis in Zea Mays L., Ph.D. thesis, U. of Wageningen Press, The Netherlands (1988).
M. Fromm et al., "Expression of Genes Transferred into Monocot and Dicot Plant Cells by Electroporation", PNAS, 82:5824-5828 (1985).
Gen. Eng. News, news article entitled "Shotgunning DNA into Cells" (Jul./Aug. 1987).
O. Gould et al., "Shoot Tip Culture as a Potential Transformation System", Beltwide Cotton Production Research Conferences, New Orleans, La., Abstract p. 91 (Jan. 3-8, 1988).
A. Graves et al., "The Transformation of Zea Mays Seedlings with Agrobacterium tumefaciens", Plant Mol. Biol., 7:43-50 (1986).
C. Green et al., "Plant Regeneration from Tissue Cultures of Maize", Crop Science, 15:417-421 (1975).
C. Green et al., "Somatic Cell Genetic Systems in Corn", Advances in Gene Technology: Molecular Genetics of Plants and Animals, Academic Press, Inc., 147-157 (1983).
P. Hooykaas, "Transformation of plant cells via Agrobacterium", Plant Mol.

Biol., 13:327-336 (1989).
M. Horn et al., "Transgenic Plants of Orchard Grass (Dactylis glomerata L.) from Protoplasts", Chem. Abstracts, 110:208, Abstract 89869A (1989).
R. Jefferson et al., "-Glucuronidase from Escherichia coli as a Gene-Fusion Marker", PNAS, 83:8447-8451 (1986).
R. Jefferson, "Assaying Chimeric Genes in Plants: The GUS Gene Fusion System", Plant Mol. Biol. Reporter, 5:387-405 (1987).
K. Kamo et al., "Establishment and Characterization of Long-Term Embryogenic Maize Callus and Cell Suspension Cultures", Plant Science, 45:111-117 (1986).
K. Kartha et al., "Transient Expression of Chloramphenicol Acetyltransferase (CAT) Gene in Barley Cell Cultures and Immature Embryos Through Microprojectile Bombardment", Plant Cell Reports, 8:429-432 (1989).
T. Klein et al., "Regulation of anthocyanin biosynthetic genes introduced into intact maize tissues by microprojectiles", PNAS, 86:6682-6685 (1989).
T. Klein et al., "Genetic Transformation of Maize Cells by Particle Bombardment and the Influence of Methylation on Foreign-Gene Expression", Gene Manipulation in Plant Improvement II, Plenum Press, N.Y., pp. 265-266 (1990).
M. Kozak, "Point Mutations Define a Sequence Flanking the AUG Initiator Codon that Modulates Translation by Eukaryotic Ribosomes", Cell, 44:283-292 (1986).
P. Lazzeri et al., "In Vitro Genetic Manipulation of Cereals and Grasses", Ad. Cell Culture, 6:291-293 (1988).
H. Lorz et al., "Advances in Tissue Culture and Progress Towards Genetic Transformation of Cereals", Plant Breeding, 100:1-25 (1988).
S. Ludwig et al., "High Frequency Callus Formation from Maize Protoplasts", Theor. Appl. Genet., 71:344-350 (1985).
S. Ludwig et al., "Lc, A Member of the Maize R Gene Family Responsible for Tissue-Specific Anthocyanin Production, Encodes a Protein Similar to Transcriptional Activators and Contains the myc-Homology Region", PNAS, 86:7092-7096 (1989).
S. Ludwig et al., "A Regulatory Gene as a Novel Visible Marker for Maize Transformation", Science, 247:449-450 (Jan. 26, 1990).
S. Ludwig et al., "Maize R Gene Family: Tissue-Specific Helix-Loop-Helix Proteins", Cell, 62:849-851 (Sep. 7, 1990).
H. Lutcke et al., "Selection of AUG Initiation Condoms Differs in Plants and Animals", EMBO J., 6:43-48 (1987).
C. McDaniel et al., "Cell-Lineage Patterns in the Shoot Apical Meristem of the Germinating Maize Embryo", Planta, 175:13-22 (1988).
M. Meadows, "Characterization of Cells and Protoplasts of the B73 Maize Cell Line", Plant Sci. Letters, 28:337-348 (1982/83).
R. Mendel et al., "Delivery of Foreign Genes to Intact Barley Cells by High-Velocity Microprojectiles", Theor. Appl. Genet., 78:31-34 (1989).

T. Murakami et al., "The Bialaphos Biosynthetic Genes of Streptomyces hygroscopicus: Molecular Cloning and Characterization of the Gene Clusters", Mol. Gen. Genet., 205:42-50 (1986).
J. Odell et al., "Identification of DNA Sequences Required for Activity of the Cauliflower Mosaic Virus 35S Promoter", Nature, 313:810-811 (1985).
P. Ozias-Akins et al., "Progress and Limitations in the Culture of Cereal Protoplasts", Trends in Biotechnology, 2:119-123 (1984).
P. Ozias-Akins et al., "In Vitro Regeneration and Genetic Manipulation of Grasses", Physiologia Plantarum, 73:565-569 (1988).
J. Poehlman, Breeding Field Crops, 3rd Ed., AVI Publishing Co., pp. 469-481 (1987).
I. Potrykus et al., "Callus Formation from Stem Protoplasts of Corn (Zea mays L.)", Molec. gen. Genet., 156:347-350 (1977).
Sanford, "Biolistic Plant Transformation", Physiol. Plantarum, 79:206-209 (1990).
A. Schmidt et al., "Media and Environmental Effects on Phenolics Production from Tobacco Cell Cultures", Chem Abstracts, 110:514-515, Abstract 230156 (1989).
R. Smith et al., "Shoot Apex Explant for Transformation", Plant Physiology (Suppl.), 86:108, Abstract 646 (1988).
Spencer et al., Poster Presentation, Faseb Plant Gene Expression Conference, Copper Mountain, Colo. (Aug. 6-11, 1989).
Spencer et al., "Bialaphos Selection of Stable Transformations from Maize Cell Culture", Theor. Appl. Genet., 79:625-631 (May 1990).
C. Thompson et al., "Characterization of the Herbicide-Resistance Gene bar from Streptomyces hygroscopicus", EMBO J., 6:2519-2523 (1987).
Tomes et al., "Transgenic Tobacco Plants and Their Progeny Derived by Microprojectile bombardment of Tobacco Leaves", Plant Mol. Biol., 14:261-268 (Feb. 1990).
D. Twell et al., "Transient Expression of Chimeric Genes Delivered into Pollen by Microprojectile Bombardment", Plant Physiol., 91:1271-1274 (1989).
E. Ulian et al., "Transformation of Plants via the Shoot Apex", In Vitro Cellul & Dev. Biol., 9:951-954 (1988).
V. Vasil et al., "Improved Efficiency of Somatic Embryogenesis and Plant Regeneration in Tissue Cultures of Maize (Zea mays L.)", Theor. Appl. Genet., 66:285-289 (1983).
V. Vasil et al., "Plant Regeneration from Friable Embryogenic Callus and Cell Suspension Cultures of Zea mays L.", J. Plant Physiol., 124:399-408 (1986).
V. Walbot et al., "Molecular Genetics of Corn", Ag. Mono., 18:389-430 (1988).
Y. Wang et al., "Transient Expression of Foreign Genes in Rice, Wheat and Soybean Cells Following Particle Bombardment", Plant Mol. Biol., 11:433-439 (1988).

J. White et al., "A Cassette Containing the Bar Gene of Streptomyces hygroscopicus: A Selectable Marker for Plant Transformation", Nucleic Acids Res., 18:1062 (1989).
I. K. Vasil, Bio/Technology, 8, 797 (Sep. 1990).

Primary Examiner: Benzion; Gary
Attorney, Agent or Firm: Schwegman, Lundberg & Woessner

Parent Case Text

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part of U.S. patent application Ser. No. 07/467,983, filed Jan. 22, 1990 now abandoned.

Claims

What is claimed is:

1. A fertile transgenic Zea mays plant of the R0 generation containing heterologous DNA encoding Bacillus thuringiensis endotoxin, wherein said DNA is expressed so that the plant exhibits resistance to an insect, wherein said expression is not present in said plant not containing said DNA, and wherein said DNA is transmitted through a complete normal sexual cycle of the R0 plant to the R1 generation, and wherein said

DNA is introduced into said plant by microprojectile bombardment of Zea mays callus cells.

2. The transgenic plant of claim 1 wherein said DNA comprises a promoter.

3. The transgenic plant of claim 1 which is selected from the group consisting of field corn, popcorn, sweet corn, flint corn and dent corn.

4. A seed produced by the transgenic plant of claim 1 which comprises a replication of said heterologous DNA.

5. An R1 transgenic Zea mays plant derived from the plant of claim 1 wherein said R1 plant expresses said heterologous DNA so that the R1 plant exhibits said phenotypic characteristics.

6. A progeny transgenic Zea mays plant derived from the plant of claim 5 wherein said progeny plant expresses said heterologous DNA so that the progeny plant exhibits said phenotypic characteristics.

Appendix C: Summary of Tables and their Component Data Fields in the Relational Agbiotech Patent and Bioscience Paper Database

Table Name	*Description and component fields*
ABSTITLE	**Patent title and abstract:** 1. Patent number 2. Patent title 3. Patent abstract
AUTH INST	**Author institutions listed on paper:** (note several addresses listed on a paper may resolve to a single institution) 1. Paper locator 2. Position (one for each institution listed in paper) 3. Author institution name 4. Count (number of mentions of this institution for this paper--may be several addresses) 5. Total count of addresses for this paper 6. Fractional count of this institution for paper (<=1)
ASSIGNEES	**Assignees for each patent:** 1. Patent 2. Position in list of assignees for that patent 3. Assignee name 4. Name for corporate parent
CITED PATENTS	**Patents cited by patents in the main set:** 1. Patent cited by at least one patent in main set 2. Application date of cited patent 3. Issue date of cited patent
CITED PATENTS PAIRS	**Patents in the main set and their cited patents:** 1. Patent in main set 2. Patent cited by this patent
CITING PATENTS	**Patents citing patents in the main set (note: both US and EPO patents):** 1. Patent citing at least one patent in main set 2. Application date of citing patent 3. Issue date of citing patent
CITING PATENTS PAIRS	**Patents in the main set and their citing patents (note: both US and EPO patents):** 1. Patent in main set 2. Patent citing this patent
FOREIGNREFS	**Foreign patents referenced (not EPO):** 1. Patent in the main set (referencing patent) 2. Cited patent (foreign)

	3. Month 4. Year 5. Country
IND	**Index between each patent and each paper:** 1. CHI's paper locator 2. Patent number 3. Position of nonpatent reference within patent 4. Position of subreference within non-patent reference (almost always 1)
INVENTORS	**Inventor names of each patent:** 1. Patent 2. Position of inventor in list for this patent 3. Inventor name 4. Inventor state 5. Inventor country abbreviation 6. Inventor city
IPC	**International Patent Classification of the main patent set patents:** 1. Patent number 2. Position in patent 3. IPC class
NPR	**Nonpatent references:** 1. Patent number 2. Number of nonpatent reference within patent 3. Text of nonpatent reference
PAP	**Each paper's information:** 1. CHI paper locator 2. Journal abbreviation 3. Journal number 4. Publication year 5. Author abbreviation 6. Page 7. Volume 8. CHI internal match status
PATENTS	**Each patent's information:** 1. Patent number (in text format) 2. Application date 3. Issue date 4. Number of cites received 5. Number of references to previously-issued US and EPO patents 6. Number of other references (non-patent references) 7. Number of claims

PATENT COUNTRY KEY	**Translation table matches three-digit alphanumeric country code to country name and to unified country code (e.g. GB6 to Scotland to United Kingdom):** 1. Alpha country code 2. Country name 3. Unified country name
POC	**US Patent Office Classification:** 1. Patent Number 2. Position in patent 3. POC class

Appendix D: Research Inputs in Other Public Sectors

Appendix D1 Research Inputs at Federally Financed R&D Centers

Federally Financed Research and Development Centers (FFRDCs) are R&D-performing entities that are exclusively or substantially financed by the federal government and are formed to meet a particular federal R&D objective that cannot be met effectively by existing organizational resources. Each center is administered either by a university or college, an industrial firm, or other nonprofit institution.

Data on research inputs at the FFRDCs are available from NSF annual report series such as "Academic Research and Development Expenditures," "Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions," and "Federal Funds for Research and Development," which are based on NSF surveys on research and development funding and expenditures.

Annual life science R&D expenditures at individual FFRDCs administered by universities or colleges are reported from 1973 to 1998. But data broken down to the disciplinary level are not collected for FFRDCs administered by industrial firms or other nonprofit institutions.

Annual data on federal obligations for research and development and R&D plant, from each obligating federal agency (e.g., USDA, NIH, and NSF), are available for individual FFRDCs from 1973 to 1998. FFRDCs' research dollars come primarily from the federal government. Hence, the federal obligation measures give us a good idea of the total research funds available at FFRDCs, taking account of the lag structure between funding obligation and real spending.

Some FFRDCs are fairly diversified. Research resulting in the outputs recorded in our patent and scientific paper datasets are only a part of the research efforts focusing on the life sciences. Thus, the research inputs measured by federal obligations for R&D do not exactly match our research output measures at given FFRDCs. The U.S. Department of Agriculture and the U.S. Department of Health and Human Services (parent agency of National Institute of Health) are the two largest sources of federal funds for research in life sciences. For example, in 1998, these two agencies accounted for 7.7% and 76.7%, respectively, of total federal obligations in life science research. All other agencies, including the Department of Defense and National Science Foundation, accounted for the remaining federal obligations. Moreover, the USDA, USHHS, and Department of Veterans Affairs are engaged primarily in providing funds for research in life sciences. [1]

In 1998, obligations for research in life sciences were 77.8%, 86.5%, and 79.6%, respectively, of the total obligations from these three agencies. All other agencies allocated much smaller proportions of their obligations to life science research, for example, 11.0%, 6.8%, and 16.0% respectively from the Department of Defense (DoD), Department of Energy (DoE), and National Science Foundation. Annual obligations for R&D in life sciences at an FFRDC can be computed as the sum of the annual federal obligations to that FFRDC from each obligating federal government agency, weighted by the ratio of obligations for life science research (divided into agricultural, biological, and medical sciences) to total obligations for

[1] The Department of Veterans Affairs is not a large funding source for research in life sciences because its total obligations for research are small compared to other federal agencies.

research in all fields from those individual obligating agencies in that year. This approach assumes the proportions do not differ across institutions in a given year. Suppose, for example, that USHHS and DoE each obligated 19.8 and 4.2 million dollars, respectively, to a FFRED in 1998. R&D obligations in the life sciences to this FFRDC are 17.4 million dollars, the sum of (19.8) (0. 865) and (4.2) (6.8). The proportions are available, on an annual basis between 1970 and 2000 and for each individual obligating agency, from the NSF report, "Federal Funds Survey, Fields of Science and Engineering Research Historical Tables."

Appendix D2 Research Inputs at Nonprofit Institutions

Nonprofit institutions are legal entities other than universities and colleges, privately organized or chartered to serve the public interest and exempt from most forms of federal taxation. This sector includes three types of institution: research institute, voluntary hospital, and other independent nonprofit institution such as professional society and private foundation.

According to reports based on the Survey of R&D Funding and Performance by Nonprofit Organizations, conducted by the NSF in 1973, 1996, and 1997, nonprofit organizations finance their own R&D activities as well as provide R&D funds to other sectors, including universities and colleges, FFRDCs, and industrial firms. At the same time, they also receive R&D funding from federal, state, and local governments, universities and colleges, and industry to perform intramural R&D within their organizations. For example, in 1997, nonprofit organizations provided $2.606 billion for R&D to all U.S. recipients, 52.4%, 0.1%, 8.9%, 15.8%, and 22.9% of which were respectively allocated to universities and colleges, FFRDCs, industrial firms, nonprofit organizations themselves, and other institution types. In that same year, the same nonprofit institutions expended \$7.349 billion in their own intramural R&D, with 50.5%, 2.4%, 0.7%, 11.2%, 5.6%, and 29.7% of this amount coming respectively from the federal government, state and local governments, universities and colleges, industry, nonprofit institutions themselves, and other sources (such as gifts and grants from private individuals and foreign sources). Hence, neither the annual total R&D funds provided by an individual nonprofit institution nor the annual total R&D funds allocated to that institution from all federal agencies can accurately represent the total intramural R&D expenditures at that institution in a given year. The NSF survey also shows that R&D conducted by nonprofit organizations is mainly in the life sciences field. For example, \$5.289 of the \$7.349 billion in total intramural R&D expenditures at nonprofit institutions was spent in the life sciences in 1997.

Although annual data on the intramural R&D expenditures at individual nonprofit institutions were collected in the NSF survey in 1996 and 1997, the time-series required for the present study are not available. The only time series available on research inputs at nonprofit institutions is that for federal obligations to individual nonprofit institutions from each federal government agency for research and development and R&D plant. These have been published on an annual basis, since 1963, in the NSF report series "Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions." Based on the 1997 survey findings discussed in the previous paragraph, R&D funds from federal agencies account for an average of only one-half of intramural R&D expenditures at nonprofit organizations. Assuming this proportion to be fixed across years, annual federal obligations completely capture the time-wise variation in total obligations for intramural R&D at these nonprofit institutions.

As most nonprofit organizations, such as voluntary hospitals and medical research institutes, concentrate on life science research, one approach to computing the total life science R&D expenditure at a nonprofit institution is to compute the sum of life science obligations to this institution from all federal agencies. In another approach, using the method discussed in section 5.2.1, annual federal obligations for R&D in the life sciences, specifically in the agricultural, biological, and medical sciences, at a nonprofit institution can be calculated based on the annual federal R&D obligations received by that institution from each obligating agency, and on the proportions of obligations for life science research made by each obligating agency in that year.

Appendix D3 Research Inputs at Government Agencies

In addition to R&D funds provided to other sectors such as universities, FFRDCs, and nonprofit institutions, federal agencies obligate funds for R&D performed directly by federal personnel, that is for intramural R&D. In some federal agencies, obligations for intramural R&D account for a significant part of total R&D obligations. In 1998, for example, 69.1%, 20.5%, and 44.6% of total R&D obligations in the USDA, USHHS, and NSF were for intramural performance.

Variables describing inputs to intramural R&D at federal agencies are reported annually by the NSF in "Federal Funds for Research and Development." Annual data on federal obligations from subdivisions of federal agencies (e.g., Agricultural Research Service, the USDA's Forest Service, and NIH and FDA within USHHS) for intramural R&D performance, and total basic and applied research expenditures, are available from 1973 to 1998. Obligations reported in this category are for activities performed, or about to be performed, by the reporting subdivision or agency itself; or they represent funds the agency transfers to another federal agency for R&D performance. As the federal government is the primary and nearly the only funding source for R&D activities carried on by federal agencies, and the amounts of funds transferred among federal agencies for intramural use are not large, these obligation measures give us fairly complete coverage of the intramural R&D funds used by such agencies.

Assuming the proportions of obligations for life science research from different subdivisions of a given federal agency equal the proportions at the entire agency in a given year, the obligations for intramural life science R&D at an individual subdivision can be calculated by multiplying these proportions by the obligations for total intramural R&D. For example, the USDA's Forest Service obligated $170.2 million for intramural R&D in 1998. Just over three-quarters (77.8%) of total research obligations at USDA were for life sciences in that year. Thus, intramural life science R&D obligations at the Forest Service are computed as ($170.2) (0.778) = $132.4 million.

Appendix E: Production Outputs, Inputs, and Other Characteristics at Private Firms

Data on production outputs and inputs at agricultural biotechnology firms identified in our patent and paper datasets are available mainly in the Longitudinal Research Database (LRD), maintained by the Census Bureau's Center for Economic Studies (CES) and linked to the R&D database. The LRD consists of manufacturing establishments' geographic locations, ownership status, cost, output volumes, input quantities (labor, materials, capital, and energy), and other variables as enumerated in the Census of Manufactures and Annual Survey of Manufactures. A supplementary source is the Bureau of Labor Statistics (BLS), which provides price indexes at aggregate SIC industry levels.

The Census of Manufactures (CM) is an enumeration of all establishments whose primary activity is manufacturing (McGuckin and Pascoe). The CM was conducted in 1963, 1967, 1972, 1977, 1982, 1987, 1992, and 1997. The Annual Survey of Manufactures (ASM) is instead a sample of establishments drawn from the CM universe. A new sample is selected after each census and is used for the next five years. Another difference between the CM and ASM is in the data items collected. In addition to the basic economic activity measures in both the CM and ASM, each contains information the other does not. For example, individual materials consumption, value of shipments and physical outputs at the 7-digit SIC level, and output and input prices are contained only in the CM. Rental payments and capital assets are contained only in the ASM. Observations on variables available only in census years may need to be interpolated in intervening years, or might be obtained from other sources, including the BLS. The linked CM and ASM form an unbalanced longitudinal database on which the LRD is based.

The Census Bureau permitted us to access, for individual biotech firms and at an establishment-level basis, the LRD raw data on outputs, inputs, and other characteristics in years 1972 to 2000, as detailed below. A number of calculations can be performed on the raw data to obtain series useful for econometric analysis.

Appendix E1 Outputs

Value of shipments and quantity of physical outputs produced and shipped are reported at the SIC 7-digit level in CM years, whereas shipment values are available for each 5-digit SIC product class in ASM years. In census years, I can calculate the shipment price index in the form of unit values by dividing total shipment quantity into total shipment value. However, quantity units are not adjusted for quality. For example, the 7-digit SIC level does not distinguish between 100 pounds of concentrated nitrogen solution and the same amount of unconcentrated one. Thus, the price indexes obtained in this way represent an "average" of all the establishment's outputs. Observations in non-census years need to be interpolated.

The Bureau of Labor Statistics (BLS) independently reports producer price indexes based on quality-adjusted products at various aggregation levels, such as in the 3-, 4-, and 5-digit SIC classes. Dividing shipment value at each 5-digit SIC level in ASM years by the corresponding price indexes gives us quantity of output in that 5-digit class. The total quantity of outputs is the sum of all SIC 5-digit output quantities.

Related information, such as value of resales, receipts for contract work, and beginning- and end-of-year inventory value, are also available from the LRD for each establishment in each year. Strictly speaking, inter-year changes in inventory and resale values should be taken into account when using shipments to represent total output. However, inventory and resale changes across years are very small compared to value-of-shipment magnitudes. Inventory and resale values are therefore ignored in this study.

Appendix E2 Inputs

In the LRD, production factors are grouped into four categories: labor, materials, capital, and energy. As an establishment is the reporting unit in the CM and ASM and hence the observation unit in the LRD, establishment's inputs are not allocable to its products. Input variables are reported at the establishment level.

Appendix E2.1 Labor

The LRD contains the following variables on labor input: number of employees, number of production workers, number of production worker hours, total salaries and wages, production worker wages, and non-production worker wages. Total labor quantity is the sum of production and non-production worker hours. Assuming each non-production employee works 2000 hours per year, the latter is the product of 2000 hours per year and the number of non-production workers, which is total employment less the number of production workers. Dividing total labor cost (payroll) by total labor quantity gives the weighted-average wage rate series.

Appendix E2.2 Materials

The total cost of materials is included in ASM years, and the delivered cost and quantity received and consumed of individual materials are available in CM years. In census years, I obtain the price for materials by dividing total quantity of materials into total material cost. Similar to the output price index, material price calculated in this way is an "average" value of all materials used at one establishment.

Material price and quantity can be computed following another approach. BLS creates a material deflator for individual SIC 4-digit industries by averaging price deflators for more than 500 inputs (about 350 manufactured inputs and 150 non-manufactured inputs), using the relative size of each industry's purchases of that input as weights. Producer price indexes in various product categories in manufacturing and non-manufacturing are used as the price deflators. The material price at each establishment can be developed as the weighted material deflators at the relevant SIC 4-digit level according to the proportionate share of each 4-digit industry in total value of shipments at that establishment. By dividing the total cost of materials by the material price, I obtain the quantity of materials used at the given establishment.

Appendix E2.3 Capital

Data on capital inputs (buildings and equipment) reported in the LRD include assets, new capital investment, rent, depreciation, retirements, and repair cost of building and machinery. These data, however, cannot be used as capital expenditures in production analysis. What is needed is the value of service provided by the capital stock, which is the rental cost of capital.

By combining separate asset deflators for structures and equipment based on the distribution of each asset type in the industry, the National Bureau of Economics Research's (NBER) manufacturing productivity database provides the acquisition price deflator for new capital spending for each SIC 4-digit industry. The rental price of capital can be computed using the following formula:

$$P_K = q_{t-1} r_t + q_t \delta_t - \left(q_t - q_{t-1} \right) \tag{E1}$$

where q is acquisition price of capital, r is opportunity cost of capital, δ is capital depreciation rate, and $(q_t - q_{t-1})$ is capital gain. The long-term bond yield can be used for r_t, and the capital depreciation rate δ can be assumed to be 1/10. The only difference between equation (E1) and the capital rental price formula in Christensen and Jorgenson (1969) is that (E1) ignores property tax, the structure of which is complicated and for which no information is readily available. The capital rental price at each establishment can then be obtained by averaging the rental prices at the SIC 4-digit level according to the proportionate share of each 4-digit industry in total shipment value at that establishment. The product of the capital stock and capital rent price is the rental cost or service flow of capital.

Appendix E2.4 Energy

Cost and quantity of purchased electricity and cost of other fuels (e.g., residual fuel oil, distillates, coal, coke, and natural gas) are available for each year in the LRD.

Based on information from various sources, such as the National Energy Accounts and the Department of Energy's State Energy Price and Expenditure Report, the price deflator for energy input at the SIC 4-digit level is provided in the NBER's manufacturing productivity database. The energy price at individual establishments can be approximated by the average of the energy deflators at the SIC 4-digit level, weighted by the proportionate share of each 4-digit industry in the total value of shipments at each establishment. Total energy cost is the sum of cost of purchased electricity and other fuels. Diving the total cost of energy by the energy price gives the energy quantity.

Appendix E3 Other Characteristics in Production

Apart from data on establishments' outputs and labor, materials, capital, and energy inputs, the LRD also provides information on the following characteristics of manufacturing establishments:

(a) Plant identification number

(b) Employer identification number

(c) Business name

(d) Geographic location (i.e., mailing address and state and county codes)

(e) Industrial classification (SIC)

(f) Legal form of organization

(g) Parent firm

(h) Status of establishment

Items (a) and (d) are permanent in the sense that they stay with the plant from its birth until it shuts down. In addition, items (g) and (h) allow one to trace an establishment's ownership changes over time.