The Influence of R&D Expenditures on New Firm Formation and Economic Growth

by

BJK Associates
Maplewood, New Jersey

for

SBA Office of Advocacy

under contract number SBAHQ-00-M-0491

Release Date: October 2002

Table of Contents

LIST OF FIGURES

THE INFLUENCE OF R&D EXPENDITURES ON NEW FIRM FORMATION AND ECONOMIC GROWTH

ABSTRACT

This paper describes research designed to determine whether university R&D activity affects the local rate of new firm formations and economic growth. We created a file of university R&D expenditures by Labor Market Area (LMA) in the U.S. and combined this with data on new business formations by LMA. The hypothesized relationships were tested using multiple regression analysis, while controlling for other relevant exogenous socio-economic variables The results show that university R&D expenditures are significantly related to new firm formations in the same LMA.

In addition, we tested for a relationship between R&D expenditures and local economic growth, measured as employment growth by LMA. The results show that university R&D expenditures are not significantly related to economic growth, once one has controlled for the birth rate in the previous period. However, the variations in the birth rates, which are affected by R&D spending, are strongly associated with the growth rates at the LMA level.

These findings lend strength to the argument that government and private sector R&D expenditures made through research universities contribute to economic growth. Although this argument has traditionally been made with the expectation of long term lags in the R&D to growth relationship, our findings are that this lag is relatively small – as little as one year and the effect seems to decrease slowly, but steadily, after the first year but lasts for at least five years. University R&D spending is also associated with localities with higher levels of human capital, which also contributes substantially toward generating new firms., Thus, research universities and investment in R&D at these universities are major factors contributing to economic growth in the labor market areas in which the universities are situated.

INTRODUCTION

The relationship between innovation and economic growth has been the topic of numerous theoretical and empirical research efforts over the past decades. Beginning with Schumpeter's (1934) classic theoretical work, the notion that innovation or the commercialization of invention, leads to economic growth has continued to gain acceptance among scholars. Specifically, Schumpeter argues that newly formed independently owned firms commercialize inventions that increase overall demand thereby causing economic growth, destroying the existing market structures, and redistributing wealth among the remaining firms in the market. Schumpeter calls the process "creative destruction."

Small Firms' Superior Innovation Ability

In technology intensive industries, inventions often originate with research and development. Private firms in capitalist economies have long believed in research and development as an important process to produce new inventions to be commercialized. This is done with the hope that the invention will become an innovation that will yield a competitive advantage and thus firm growth. However, Scherer (1984) notes that scale economies do not seem to benefit large firms when it comes to innovation. In fact, from his analysis of 448 firms on the 1955 Fortune 500 list, he concludes that although large firms may spend more than small firms on research and development (R&D), the process of creating inventions that are the foundation for innovation, there is "little evidence of disproportionately great R&D input or output associated with the largest corporations" (p. 170). Additionally, in a report prepared for the U.S. Small Business Administration, The Futures Group (1984), replicating an earlier study by Gellman (1976), analyzed a database that consists of 8,074 product innovations introduced in the US market during 1982. The Futures Group finds evidence that small firms actually innovate at a rate of 1.24 to 2.38 times that of large firms. Furthermore, in his analyses of the Gellman (1976) product innovation database, Audretsch (1991, 1995) initially concludes that small firms are not necessarily at an innovative disadvantage in capital-intensive industries, and later adds that the ability to innovate may allow small firms to offset the scale disadvantages relative to large firms and may ultimately lead to greater growth and survival rates. Lastly, in a study of small and medium sized U.S. firms, Chakrabarti (1991) finds that small firms actually generate more innovations per R&D dollar than large firms. In summary, it seems a paradox exists; how are small firms able to innovate so successfully despite their limited investment in R&D?

Spillover Effects

A major link has been made that explains this paradox. The answer to this question can be found in studies examining the relationship between new technical knowledge and the spatial distribution of innovative activity (Acs, Audretsch and Feldman, 1994; Anselin, Varga and Acs, 1997; Audretsch & Feldman, 1996; and Jaffe, 1986, 1989). Jaffe finds empirical evidence to suggest that the tacit knowledge generated by the R&D process at one firm may spill over to and be subsequently exploited for economic gain by other firms. Moreover, in their retrospective analyses of the Gellman (1976) database, Acs et al. (1994) conclude that this flow of knowledge is unidirectional (specifically, that knowledge seems to flow from large to small firms) and that geographic proximity acts as a catalyst for this knowledge transfer (especially for small firms). Audretsch & Feldman (1996) add that these spillovers lead to innovative clustering and are most prevalent in R&D intensive industries. These findings suggest that while small firms may invest less in R&D than large firms, they are able to gain access to much of the same new knowledge upon which "innovation and technological change depend" as their larger counterparts through knowledge spillovers, especially in those economic sectors in which this knowledge is considered particularly valuable (Audretsch & Feldman, 1996: 630).

Small Firms and Economic Growth

This knowledge spillover effect provides a logical rationale for the empirical research findings about the role of small, and especially newly formed small firms. Many researchers, in a variety of different nations, have produced empirical research that shows small firms make a significant contribution to economic growth as measured by net new job creation (Birch, 1979, 1987; Kirchhoff, 1994; Storey, 1994; Baldwin, 1995; Wennekers and Thurik, 1999; Stel, et al., 2002; Almus and Nerlinger, 1999). Moreover, among small firms, newly formed firms create the largest share of net new jobs (Kirchhoff, 1994; Baldwin, 1995; Wennekers and Thurik, 1999). And, among newly formed firms, highly innovative new firms create a disproportionately greater share of net new jobs than those new firms with lesser innovation intensity (Kirchhoff, 1994). This suggests that highly innovative new firms are a major source of economic growth. This last sentence is, of course, Schumpeter's (1934) original hypothesis.

University R&D and New Firm Formations

But spillover effects need not be limited to corporate R&D spillover to new small firms. University research laboratories are equally likely to exhibit the same spillover effects. To date, the research on such effects have focused on spin-off new firms typically started by one or more faculty from university R&D labs. Business incubators have been constructed at many research universities to accommodate such spin-offs. Tesfaye (1997) reviews the extensive literature on university spin-offs in Europe and North America. He then identifies and describes 21 successful new firms in the Stockholm area that spun-off from Stockholm University and the Royal Institute of Technology. Given the existence of university R&D spillovers, one can expect that new firm births would be correlated with the extent of R&D activity at research universities. As with private firm R&D, such effects would appear as clustering in the local area of the university R&D activity.

The above logic leads to two research questions: (1) Do R&D activities at research universities have a significant effect on local new firm formations? And: (2) do R&D activities at research universities have a significant effect on local economic growth? It is these questions that we seek to answer with the research reported here.

HYPOTHESES

Previous Research

Two major prior studies have attempted to explain the possible link between regional innovative activity, or knowledge, and new firm formation. The first was conducted by Reynolds, Miller and Maki (1995). In their research on new firm births and deaths in the U.S., the authors identify fifteen factors that may influence the rate of new firm formation, one of which was "access to research and development, information, and innovation." The authors hypothesize that "where information is readily available and innovation and creativity flourish, the formation rate of new firms is enhanced" (Reynolds et al. 1995: 391). While they find no evidence to support this hypothesis, Reynolds et al. (1995) acknowledge that this lack of statistical support may have been in part due to invalid measurement. Specifically, they suggest that their operationalization of the independent knowledge/innovation variable (quantified as the density of post-college adults, professional and technical employees, patents granted, or doctorates earned in a given metropolitan area) may not have been appropriate.

The second was conducted by Armington & Acs (2001). They find some evidence in support of Reynolds et al.'s (1995) hypothesis in their analysis of the factors underlying new firm formation using the Longitudinal Establishment and Enterprise Microdata (LEEM). Armington & Acs (2001) conclude that firms are more likely to form in labor market areas (LMAs) that have a high percentage of college graduates than in those LMAs with high concentrations of less skilled workers, suggesting that a positive relationship may actually exist between the "size" of a region's knowledge base and new firm formation rate.

There is little other research on this topic that deals specifically with local effects, in great part because of the earlier lack of appropriate data. The relatively recent development of the "local" measure defined as Labor Market Areas has provided a basis for aggregating county level data to construct local economic units. LMAs within the U.S. (50 states plus DC) are defined according to the 1990 specification of Tolbert and Sizer (1996) for the Department of Agriculture. There are 394 LMA's, all based on aggregations of counties, many of them cutting across state boundaries. We use the most recently specified state and county for each establishment in the LEEM, assuming that most of the few location coding changes are corrections. Businesses that report operating statewide (county =999) have been placed into the largest LMA in each state. LMAs are defined not only by the place of work but also by where the workers live – i.e., commuting routes linking work to workers (Tolbert and Sizer, 1996).

The research by Reynolds et.al. (1995) may have failed to find significance because they did not use LMAs as their regional identifiers. To overcome this difficulty, we have assembled a broad range of data based on the 394 LMAs in the U.S. For this type of research, LMAs are far superior to Metropolitan Statistical Areas (MSAs) because LMAs cover all of the U.S., not just the cities, and there are research universities in rural areas (e.g., University of Iowa at Iowa City; University of Alabama at Tuscaloosa; University of Missouri at Rolla) that are not in MSAs but are in LMAs. Furthermore, LMAs link the workers to the work place so that measures of college educated adults are truly linked to the place where they work. MSAs include areas of population density but not necessarily suburbs where most college educated adults live. This may be why Armington and Acs (2001) were able to find a relationship between college educated adults and firm births while Reynolds (1995) was not.

In summary, our literature review develops the argument that university R&D creates inventions that spill over into commercialization by new firm formations. Intuitively, it is reasonable to expect that this spill-over does not happen simultaneously with the R&D expenditures. We expect that there will be a lag between the R&D activity and firm births.

Hypotheses One and Two

Following Reynolds, et al. (1995) and Armington and Acs (2001) we hypothesize:

> *H1 LMA new firm formation rates will be positively related to (a) LMA university research and development expenditures and (b) LMA human capital.*

However, we cannot ignore other social and economic factors that affect new firm formations. These must be included in our models to control for their effect.

Based upon the literature cited above showing the linkage between firm births and economic growth, we also hypothesize:

> *H2 LMA economic growth rates will be positively related to (a) LMA new firm formation rates, and (b) LMA university research and development expenditures, and (c) LMA human capital.*

RESEARCH METHODOLOGY

There are a number of socio-economic variables measuring other important differences in the environment for entrepreneurial activity. We chose environmental factors that prior research has shown are likely to affect firm birth rates. These factors are included in our regressions to control for environmental effects so that the effect of R&D and knowledge base can be isolated and identified.

Our first step in testing these hypotheses was to create a database with the necessary "primary" variables. We define these as primary variables since many of the actual variables in the regression analysis are transformations of these variables, e.g. firm births are measured as firm births divided by number of persons in the labor force. As will be explained later in this section, such ratios have advantages over the primary numbers. In the following paragraphs, we will cite the literature on economic growth and firm births as a basis for selecting the exogenous variables that are most likely to be additional factors affecting firm birth rates.

Primary Variables Subject to Hypothesis

Firm Births

We obtained firm births by LMA for 1990 through 1996 from a data file prepared by Armington and Acs (2001), using an earlier version of the LEEM file at the Center for Economic Studies of the Bureau of the Census. The same procedures were followed to tabulate firm births for 1997 through 1999 from more recent LEEM files. These later tabulations were prepared by the Company Statistics Division at the Census Bureau, under contract to the Kauffman Center for Entrepreneurial Leadership, which also provided the funding for the earlier work.

Some additional explanation about this database and how firm births are identified is useful here. The current LEEM file facilitates tracking employment, payroll, and firm affiliation and (employment) size for the more than eleven million establishments that existed at some time during 1989 through 1999. This file was constructed by the Bureau of the Census from its annual Statistics of U.S. Business (SUSB) files,[1] which were developed from the economic microdata underlying Census' County Business Patterns. These annual data were linked together using the Longitudinal Pointer File associated with the SUSB, which facilitates tracking establishments over time, even when they change ownership or identification numbers.

[1] The SUSB data and their Longitudinal Pointer File were constructed by Census under contract to the Office of Advocacy of the U.S. Small Business Administration. For their documentation of the SUSB files, see Armington (1998).

The basic unit of the LEEM data is a business establishment (location or plant). An establishment is a single physical location where business is conducted or where services or industrial operations are performed. The microdata describe each establishment for each year of its existence in terms of its employment, annual payroll, location (state, county, and metropolitan area), primary industry, and start year. Additional data for each establishment and year identify the firm (or enterprise) to which the establishment belongs, and the total employment of that firm.

A firm (or enterprise or company) is the largest aggregation (across all industries) of business legal entities under common ownership or control. Establishments are owned by legal entities, which are typically corporations, partnerships, or sole proprietorships. Most firms are composed of only a single legal entity that operates a single establishment—their establishment data and firm data are identical, and they are referred to as "single-unit" establishments or firms. The single-unit businesses are frequently owner-operated. Only 4 percent of firms have more than one establishment, and they and their establishments are both described as multi-location or multi-unit.

Firm births include both new single-unit firms with less than 500 employees, and the primary locations of new multi-unit firms with less than 500 employees, firm-wide. Those new firms that had 500 or more employees in their first year of activity appear to be primarily offshoots of existing companies. Annually, there were somewhat fewer than 150 such large apparent births of single-unit firms, with an average of about 1500 employees each. About a third of these larger single-unit firms were employee-leasing firms or employment agencies, while the remainder were widely distributed across industries. In contrast, examination of the new firms with 100-499 employees in their first year showed that most seemed credible startups, frequently in industries that are associated with large business units, such as hotels and hospitals. Since this study is not concerned with the direct employment impact of firm births, there is little risk of bias in the aggregate birth counts as a result of inclusion of a few larger startups that might actually be offshoots of existing businesses. Therefore, the startups with 100 to 499 employees were included, if they qualified otherwise.

Single-unit firm births in year t are identified on the LEEM as non-affiliated establishments with a start-year of "t" or "t" that had no employment in March of year t, and had positive employment below 500 in March of year t. This avoids inclusion of either new firms that have not yet actually hired an employee, or firms recovering from temporary inactivity. The 'start-year' is the year that the establishment entered the Census business register. About 400,000 new firms generally appear in the business register (with some positive annual payroll) the year before they have any March employment, and we postpone their 'birth' until their first year of reported employment. An average of 90,000 older firms each year reduce their March employment to zero and then recover the following year, and they will not be included as births.

We have also included most of the relatively few multi-unit firms (1500 to 6000 per year) that appeared to start up with less than 500 employees in multiple locations in their first year. We limited multi-unit firm births to those whose employment in their new primary location constituted at least a third of their total employment in the first year.[2] This rule effectively eliminated the 600 to 1000 new firms each year which were apparently set up to manage existing locations -- relatively small new headquarters supervising large numbers of employees in mainly older branch locations which were newly acquired, or perhaps contributed by joint venture partners.

[2] We tested a similar rule using one-half and found that the primary difference was in quite small multi-unit firms, where the smaller share was credible for the first year.

University R&D Expenditures[3]

We selected R&D expenditures as the measurement of R&D activity since it is reasonable to assume that activity is proportional to expenditures, and data on expenditures was available by institution. The National Science Foundations uses an annual survey of research universities and colleges in the U.S. to collect data on separately budgeted R&D spending for science and engineering (including all sources of funding, about 60% of which is federal). They use a certainty sample of universities that award science and engineering PhD's, and those that are traditionally Black and award Masters degrees, and of all 18 federally funded R&D Centers. Prior to 1992 they also surveyed about a 25% sample of other smaller colleges and universities that have had at least $50 million of separately funded science and engineering R&D. This latter group is fully covered in the annual surveys since 1992, which cover around 600 institutions, while around 500 were covered prior to 1992. The data for 1989 to 1991 for those few not previously covered, but with large expenditures already in 1992, have been set to their reported 1992 levels, but this approximation accounts for a very small portion of the total. NSF claims that their survey accounts for 95% of total US academic science and engineering R&D expenditures. The National Science Foundation survey data include Zip codes for the university laboratories, which we converted first to their corresponding state and county codes, and then to their LMA codes.

It is worthwhile to note at this point that university R&D expenditures include both federal government and private sector funding of research. However, little of the private sector funding comes from for-profit corporations. The National Science Board (2002, p. 10) reports corporate funding of 1.5 percent of academic R&D in 1994.[4] By 1999, this had risen to 8 percent. On the other hand, 58 percent was funded by the federal government. The remaining 34 percent was funded by private foundations, state governments, and by the academic institutions themselves (National Science Board, 2002, p. 13).

SBIR and STTR Grants

We also obtained data from the National Science Foundation on the Small Business Innovation Research grants and the Small Business Technology Transfer Research grants for the years 1994 through 1999. These grants provide funds for small businesses and universities for developing scientific information useful to the federal government. Some of the STTR grant money will also be reported in the university R&D data collected by NSF, but these are very small amounts compared to the total R&D expenditures. The grants made by these two programs were combined, by adding them together, since their functions seemed to be similar, and the separate series were pretty sparse.

Employment

Employment is the most consistent measure of the size of businesses, since it is more closely related to value added than is revenue, especially when aggregating over a broad range of industries. We will use it to calculate the average size of establishments in each LMA, as a measure of the extent to which their economies are dominated by a few large businesses. Employment used here is constructed

---

[3] We had hoped to estimate private sector R&D by estimating such from the BLS data on number of scientists and engineers by LMA. Unfortunately, the researcher at BLS had incorrectly advised us that such information was available by LMA and/or county. In fact, these data are only available for MSAs, which cannot be converted to LMAs, thereby frustrating our efforts to construct estimates of private sector R&D.

[4] Unfortunately, we were not able to include corporate R&D expenditures because those data were not available on a consistent or complete basis. While some additional data on R&D expenditures of private firms are available, they only cover large firms with separately budgeted R&D. Furthermore private firm R&D expenditures are not classified by locale of expenditure but by locale of firm headquarters. For example, the greater New York City area, headquarters to a great many large corporations, would likely have a disproportionate share of corporate R&D expenditures, although most of their R&D facilities are located elsewhere. Incorporating this information into our analysis would distort the LMA R&D expenditure patterns.

from U.S. Bureau of the Census' county-level employment reported in County Business Patterns. Thus, the employment data covers the same universe of private non-farm establishments that is covered by our firm birth data.

Rate of Change in Employment

Kirchhoff (1994) and others have shown that firm births are positively associated with economic growth. Since we anticipate that firm births are related to R&D, we will examine whether R&D is also related to economic growth, beyond its contribution to higher birth rates. To measure economic growth, we chose a widely used indicator: rate of employment change. Rate of change in employment is measured as the change in employment from year t to year t+1, divided by employment in year t. See below for the specification of the employment data.

Human Capital - College Education

Glaeser (2000) cites a variety of empirical research efforts that find a strong connection between human capital and urban area growth. The most commonly used human capital measure has been the concentration of college-educated adults, although some long term studies use other measures of the level of worker skills. Unfortunately, worker skills data are not available by LMA so we use college-educated adults. The interpretation of this effect is that human capital increases the new idea production and therefore new firm births.

We obtained data on the number of adults with college degrees, and the total number of adults (population over 24 years old) in each county from the 1990 Census of Population, and aggregated both to LMA level, to calculate the share of adults with college degrees in each LMA in 1990. Unfortunately, this measure is not available for non-Census years, nor is this educational attainment measure available by county yet from the 2000 Census. Thus this indicator is not a time series, but is a constant for each LMA throughout the period of analysis. However, since a very small fraction of adult population changes annually, the relative levels of educational attainment cannot change substantially from year to year.

Human Capital - High School Education

A more basic measure of the level of education in an area is based on the proportion of the adult population that has high school degrees. We obtained number of persons over 24 with high school education by county from the 1990 Census of Population. This was converted into LMA data, and divided by each LMA's number of adults over 24 in 1990, to calculate an alternative indicator of human capital in each locality.

Human Capital - Foreign-born Population

Recent work by Reynolds (1995) has shown much higher new firm formation rates among immigrants, perhaps due to their more limited potential for employment in existing firms. Therefore we expect that LMAs with higher shares of foreign-born population will have higher firm birth rates. These data are based on annual estimates for counties from the Census Bureau, as aggregated into LMAs by Vinod Sutaria for the Kauffman Center for Entrepreneurial Leadership.

Primary Environmental Control Variables

The data on annual numbers of new firms with employees in each LMA are available only for the period from 1990 through 1999. We use all ten years of data, to maximize the reliability of the analysis, and to increase the probability of identifying important lags in the relationship of firm births to our explanatory variables. Therefore, we assembled the control variables at the LMA-level for this period also, as well as for some earlier years, to allow for testing of lags.

Population

It is intuitively appropriate to expect that the number of new firms created will be affected by the number of persons in the area. Not only is local population size needed to standardize other variables to control for the wide differences in sizes of LMAs, but also we expect a positive relationship of births to population itself, as a general indicator of agglomeration effects. We use the Bureau of the Census published data on population by county, and aggregate it into LMAs.

For 1990, these population data are based on the 1990 Census, and for other years they are estimated for each county by the Census Bureau from their annual surveys.

Population Change

The recent rate of change in the local population may be the most useful proxy for the general attractiveness of the location for both people and for businesses. Furthermore, the growth in population will contribute to the formation of new firms as a result of changes in the demand for goods and services, which tend to change in proportion to the growth in population. We use the annual average compound rate of population growth over the prior two years, calculated as the square root of the ratio of current population to the population two years earlier, minus one.

Establishment Density

Economists argue that the extent of competition is a factor that affects firm births. Here we select establishment counts to represent competition. Glaeser (2000)[5] reports several empirical research studies that used establishments and found this to be a significant factor in determining growth. The establishment count is a better measure of business activity than firm count since many local activities are carried out in establishments that are owned by firms located outside of the LMA.

Establishment counts are the number of private sector non-farm establishments with employees, tabulated from the LEEM file. This number is smaller than the number reported by Census in County Business Patterns, because we have limited our count to active establishments, while CBP includes establishments that have had any payroll during a year, even if they have no employees in the March 12 pay-period. The variable used to compare the numbers of establishments across LMAs is the number of establishments per 1000 population, using the population figures discussed above as the denominator, after dividing it by 1000.

Establishment Size

Since research shows that small firms are more innovative than large firms (Gellman, 1976; Futures Group, 1984), we need to control for the relative concentration of work activity in large firms. Areas dominated by larger establishments are also likely to have fewer new businesses because the labor force is tied up in secure long-term employment; and because larger businesses have lower innovation rates. Average establishment size is used as an estimate of this concentration. It is calculated as total employment in the LMA, divided by total number of establishments.

Unemployment Rate

Past research has found conflicting evidence about whether higher unemployment leads to more new firm births, or the contrary. The unemployment rate reflects the degree of employment distress in the LMA. High unemployment rate is normally associated with a poor economic environment. However, some theorists argue that unemployment can provide an incentive for individuals to form a business. Reynolds, et al. (1995) found very little support for this argument in his survey of the U.S. population. We will use data estimated by BLS on the number of unemployed and the labor force in each county (converted to LMAs) to calculate the annual unemployment rates for each LMA.

---

[5] Glaeser (2000) provides a literature review of the variables that can affect economic growth. Although the available literature and Glaeser's analysis focuses upon MSAs as the area variable, his review of previous economic analyses serves as the foundation for several of the variables used in this research.

Figure 1
Summary Descriptions of the Primary Variables' Sources and Data Years

Variable Name	LMA Measurement[1]	Source of Data	Data Years
Firm Births	Number of births	Census LEEM File	1990 - 1999
University R&D Expenditures	Total dollars in thousands	NSF Annual Surveys[2]	1990 – 1999
SBIR & STTR Grants	Total dollars	NSF Data[2]	1994 - 1999
Employment	Number of employees of all private non-farm businesses.	County Bus. Patterns	1990 – 1999
Rate of Change in Employment	Compound rate of growth from previous two years	Census County Business Patters	1990 – 1999
Human Capital – College Education	Share of population over 24 with college education	U. S. Census of Population	1990
Human Capital – High School Education	Share of population over 24 with high school education	U. S. Census of Population	1990
Human Capital – Foreign Born Population	Share of population born outside the U.S.	U. S. Census of Population	1990
Population	Number of residents	U. S. Census and Annual Surveys	1990 – 1999
Population Change	Compound rate of growth from previous two years.	U. S. Census and Annual Surveys	1990 – 1999
Establishments	Number of establishments tabulated from LEEM file.	Census LEEM file	1990 – 1999
Establishment Density	Establishment divided by population times 1000.	LEEM file and Census of Population	1990 – 1999
Establishment Size	Number of employees divided by number of establishments	County Bus. Patterns Census LEEM file	1990 – 1999
Unemployment Rate	Number of unemployed divided by labor force	Bureau of Labor Statistics	1990 – 1999
Labor Force	Number of employees + proprietors and unpaid family members + unemployed	Bureau of Labor Statistics	1990 - 1999
Trend	Numbers from one to ten		1990 – 1999
National Economic Growth Rate	Annual compound growth rate of U.S. gross domestic product	Bureau of Economic Analysis	1990 – 1999

1. Most of the data sources report data for counties. We used a computer software application to convert counties into LMAs.
2. NSF survey data is coded by ZIP Code. We used a computer software application to convert ZIP Codes into LMAs.

<u>Labor Force</u>

The labor force includes those people who are formally employed, those who are proprietors of their own businesses or are unpaid family members working in such businesses, and those who are currently unemployed. This is a much broader measure of the size of the local economy than we get from the employment data alone. It indicates the size of the pool of potential entrepreneurs in the area. Few people try to start new businesses in different areas than those in which they have lived and worked.

<u>Trend</u>

Use of a trend variable makes allowances for other changes over time in the relationship of firm births to R&D and the local knowledge base, or to some of the control variables, which might include an irrelevant trend in their values with the passage of time. Therefore to adjust for any such systematic, time related growth, we include a variable representing the passage of time. This is simply the number of the year of the primary independent variables less 1989, so it varies from 1 to 9.

National Economic Growth Rate

The rate of change in Gross Domestic Product for the US as a whole is used to control for changes in the national economy over time, since that should affect all LMAs, and may explain some of the differences in birth or growth rates over time.

Table 1
Description of the Primary Variables in the Birth Regression

Variable	n	Minimum	Maximum	Mean	Std. Deviation
Population	3546	100,076	15,811,237	657,023	1,270,436
Population Change	3546	-.030	.076	.009	.010
Establishments	3546	1,775	317,656	14,755	28,933
Establishment Size	3546	7.867	23.711	15.252	2.912
Establishment Density	3546	1.118	5.081	2.195	.357
Unemployment Rate	3546	.017	.303	.062	.025
College Education	3546	.069	.320	.159	.050
High School Education	3546	.459	.883	.721	.080
Foreign Population	3546	.002	.329	.032	.042
Labor Force	3546	42,293	7,866,393	333,073	650,392
R&D t	3546	0	2,106,752	50,457	151,625
Births t + 1	3546	98	30,027	1,235	2,563

Transformation of the Regression Variables

Regression analyses across local areas and over time are plagued with multiple large first order (or Pearson) correlations among several of the variables. This is especially true for most of the raw descriptive variables, because areas that are more developed will show higher levels of many variables that are associated with economic growth. Furthermore, many of the variables of interest have variances that are proportional to the size of the LMA, leading to problems with heteroskedasticity. Indeed, because the size distribution of the LMAs is skewed to the right, with many more small economic areas than large ones, and a few very large ones, the distribution of the levels of most descriptive variables are also skewed to the right.

To avoid most of these problems, we construct variables that measure the size of most explanatory factors in relation to the size of the LMAs. Then these are analyzed in terms of how each LMA varies from the average of all LMA's. Other variables, whose impact on the dependent variables is expected to be proportional to the percentage differences in their size, are converted to natural logarithms. These transformed variables in our regressions are described below, beginning with the variable names used in the following tables. Those variables not mentioned below enter the regressions without further transformations.

R&D Expenditures

For the regression analysis, R&D expenditures were divided by the labor force to avoid the implicit magnitude correlation with population. Simply stated, larger LMAs are likely to have more research universities and therefore more R&D expenditures. By dividing R&D by the labor force we produce a ratio that is adjusted for LMA size, and indicates the dollars of annual academic R&D expenditures per worker in each LMA. We took the log of this ratio so that its impact would vary in proportion to its size, rather than varying as an additive function of level. Thus an expenditure per worker of $1,500 would have 50% greater impact on births than an expenditure of $1,000 per worker. But if the same $500 difference represented LMAs with $100 versus $600 per worker, it would be associated with a 6-fold greater impact on births. Since this variable is usually lagged in our regressions we use the nomenclature “R&D t-n” for its identifier.

Grants

The Small Business Innovation Research grants and the Small Business Technology Transfer Research grants were combined, by adding them together, since both routinely include some degree of university involvement, and the separate series are thin in their geographic distribution. Moreover, the levels of these grants varied widely as a function of the size of the specific projects being funded, so a dummy variable was constructed that merely indicates the LMAs with any such grant in a given year. Among the 2,364 observations of LMAs from 1994 through 1998, there are 942 with grants and 1,422 with none.

Births t+1

A size effect similar to that with R&D influences the firm birth data. To correct for this, we multiply births by 1,000 and divide by the labor force in the previous year, so that we have this dependent variable expressed as an annual birth rate -- the number of annual births of new firms per 1000 workers in the LMA in the prior year. Note that through-out these regressions, births are considered to be the outcome of the conditions in the previous year, and they are standardized by the size of the labor force in the previous year, which is the pool of potential entrepreneurs.

Figure 2
Summary Descriptions of Transformed Variables Used in the Regressions

Variable Name	Transformation Procedure
R&D t-n	Natural logarithm of NSF data times 1000 divided by labor force. R&D always lags other data by at least one year. See explanation accompanying Table 3.
Grants	Dummy variable. Zero for no SBIR nor STTR grants in the LMA. One for one or more SBIR or STTR grants in LMA.
Births t+1	LEEM number of firm births times 1000 divided by labor force. Births are always leading other data by at least one time period.
Population log	Natural logarithm of population.
Unemployment log	Natural logarithm of unemployment rate.

Population log

The agglomeration effects of larger numbers of people in an economic area are not expected to vary directly with absolute differences in the number of people, but with proportional differences in the number of people. By using the natural logarithm of population in our model, its impact will be measured in terms of percentage differences in sizes of populations.

Unemployment log

Exploration of the relationship between births and the unemployment rate showed that the relatively few extremely high values for unemployment were not associated with proportionately extreme birth rates, so the unemployment rate was transformed to its natural logarithm. This also had the effect of producing a much more normal distribution.

HYPOTHESIS ONE: BIRTH REGRESSIONS RESULTS

H1 New firm formation rates will be positively related to (a) LMA university research and development expenditures and (b) LMA human capital.

Before the regressions can be performed, we need to check for the expected multicollinearity and to determine the appropriate lead relationship for R&D. After this, we report the results of the regression and then examine the effect of multicollinearity on this regression result.

Multi-collinearity

Since it is not uncommon to have multi-collinearity in multivariate models of time series data of economic variables, we tested for this with the Pearson correlation coefficients. These are shown in Table 2 that shows some of the variables are significantly correlated with each other. Such findings are not unusual given the similarity of the theoretical underpinnings associated with the economic variables as well as the fact that the sample contains a very large number of observations – 3,940. A careful examination reveals that in most cases the correlations are not alarmingly high, as the highly correlated relationships (i.e. over 0.40) among independent variables are limited to five coefficients. This suggests caution in interpreting results but does not represent a serious problem in modeling the relationships in a multivariate regression.

Table 2 Pearson Bivariate Correlations and Statistical Significance of Birth Model Variables

N = 3,152

	Variables	1	2	3	4	5	6	7	8	9	10
1	Population log										
2	Population Change	0.196 (0.000)									
3	Establish-ment Density	0.052 (0.002)	0.032 (0.059)								
4	Establish-ment Size	0.493 (0.000)	0.086 (0.000)	-0.294 (0.000)							
5	Unemploy-ment Rate log	-0.105 (0.000)	0.017 (0.322)	-0.419 (0.000)	-0.285 (0.000)						
6	High School Education	0.332 (0.000)	0.120 (0.000)	0.540 (0.000)	0.041 (0.015)	-0.443 (0.000)					
7	College Education	0.608 (0.000)	0.245 (0.000)	0.397 (0.000)	0.205 (0.000)	-0.383 (0.000)	0.701 (0.000)				
8	Foreign Population	0.538 (0.000)	0.297 (0.000)	0.025 (0.143)	-0.007 (0.673)	0.196 (0.000)	0.134 (0.000)	0.403 (0.000)			
9 **b**	Trend	0.026 (0.124)	-0.014 (0.390)	0.106 (0.000)	0.116 (0.000)	-0.307 (0.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)		
10	R&D t**	0.486 (0.000)	0.133 (0.000)	0.088 (0.000)	0.271 (0.000)	-0.242 (0.000)	0.352 (0.000)	0.701 (0.000)	0.233 (0.000)	0.038 (0.024)	
11	Births t + 1	0.061 (0.000)	0.486 (0.000)	0.501 (0.000)	-0.403 (0.000)	0.014 (0.405)	0.180 (0.000)	0.274 (0.000)	0.208 (0.000)	-0.062 (0.000)	0.075 (0.000)

** We will eventually use R&D t-1 in our regressions. But we will not repeat this table because the coefficients are almost identical.

Analysis to Determine Appropriate Lead for R&D with Births

Next we checked to see if a lag was evident between R&D t and Births t+1. Since these are significantly and positively correlated, we ran a series of simple regressions with births as the dependent variable. R&D was used with a lead of zero through four years. However, since births are lagged by one year, the first regression actually has R&D at time t, which is a one year lead before Births t+1. The results of these regressions are shown in Table 3. Note that when R&D in time t is regressed with births in t+1, the standardized regression coefficient is statistically significant: 0.076. When R&D is regressed with a two year lead (R&D t), the standardized regression coefficient is statistically significant and a positive .075. Increasing lead times for R&D continue to show statistically significant positive coefficients, but smaller with every additional year. And, the R^2 becomes smaller. With a five year lead time, t-4, the coefficient is significant at p = .01 but the R^2 falls to .004.

Table 3 Simple Regression Results
Dependent Variable = Births in Year t + 1

Variable	Standardized Coefficients				
R&D t	0.075 ***				
R&D t – 1		0.076 ***			
R&D t – 2			0.072 ***		
R&D t – 3				0.070 ***	
R&D t – 4					0.063 **
F	20.025 ***	20.692 ***	16.291 ***	13.502 ***	9.554 **
Adjusted R^2	0.005	0.006	0.005	0.005	
Number of observations.	3546	3152	2758	2364	1970

These results suggest that R&D does have a lasting, statistically significant relationship with births for up to five years, but this impact is greatest during the years immediately after the R&D infusion. Its effect dwindles with longer lead times.

Regression Results

Next we created a multiple regression model using the transformed variables in year t as independent variables and Births in year t+1 as the dependent variable. The results are shown in Table 4. Sample size is reduced to 3152. This is due to the loss of one year (394 observations), as a result of basing future business formation on current conditions as explanatory factors. Thus 1999 business formations are modeled as the result of independent variables in 1998, and similarly for each year back to 1991 firm births based on 1990 independent variables. Second, the use of R&D t-1 eliminates the year 1998 because there is no value for R&D t-1 in 1998. The combined effect is reduction of the total observations to 3,152.[6]

Note that the Model A in Table 4 shows the regression results without R&D while Model B shows the regression results with R&D included. R&D has a significant, positive beta coefficient and it adds a statistically significant change to the F statistic. Furthermore, R&D's standardized beta coefficient (.068) is larger than those for Trend (-.051), foreign population (-.061), and unemployment rate (.043). And, the addition of R&D in Model B increases the adjusted R^2 from 0.614 to 0.617.

Also, all the environment control variables are also significant and most are in the direction we projected, positive for population, population change, establishment density, and unemployment rate. As these variables increase in value, so does the number of births. Establishment size shows the expected negative relationship with area firm birth rates.. As the average establishment size increases (caused by the presence of many large establishments), the number of births decline.

[6] Just to affirm that the reduction in number of observations and the elimination of 1999 and 1998 from the calculations have no adverse effects on the results as reported here, we also ran all statistical analysis with R&D t. The statistics are slightly different but agree with relative size of the beta coefficients and their significance compared to those reported here.

Table 4 Regression Results
Dependent Variable = Births (year t+1)

Variable	Standardized Beta Coefficients	
	Model A	**Model B**
(constant)	0 ***	0 ***
Population log	0.229 ***	0.201 ***
Population Change	0.497 ***	0.494 ***
Establishment Density	0.467 ***	0.475 ***
Establishment Size	-0.402 ***	-0.401 ***
Unemployment Rate log	0.031 *	0.043 **
High School Education	-0.170 ***	-0.183 ***
Foreign Population	-0.060 ***	-0.061 ***
Trend	-0.052 ***	-0.051 ***
R&D t – 1		0.068 ***
F	704.621 ***	635.043 ***
Change in F		30.849 ***
Adjusted R^2	0.614	0.617
Number of obs.	3,152	3,152

Testing for Multicollinearity

Two of the environmental control variables, high school education and foreign population, are negative contrary to our expectations.. Earlier we suggested that the Pearson correlations for these two variables suggested possible multicollinearity (see Table 2) problems, but further analysis of these relationships failed to offer a credible explanation, or to change the signs of these variables. Thus, we believe multicollinearity may cause the relative importance of these beta coefficients and their magnitudes to be unreliable.

To examine this, we replaced high school education with college education as the proxy for LMA human capital, thereby creating the models in Table 5. Note that college education is significant with a positive regression coefficient in Model C. But, since college education is highly correlated with R&D (r = 0.701), when R&D is added to the regression for Model D, college education loses some of its significance, while taking away some of the influence of R&D, and making R&D no longer significant in Model D.

Table 5 Regression Results
Dependent Variable = Births (year t + 1)

Variable	Standardized Beta Coefficient	
	Model C	**Model D**
Constant	0 ***	0 ***
Population log	0.149 ***	0.148 ***
Population Change	0.481 ***	0.482 ***
Establishment Density	0.388 ***	0.393 ***
Establishment Size	-0.387 ***	-0.387 ***
Unemployment Rate log	0.104 ***	0.103 ***
College Education	0.059 ***	0.042 *
Foreign Population	-0.071 ***	-0.069 ***
Trend	-0.022	-0.023 *
R&D t - 1		0.021
F	663.448 ***	590.044 ***
Change in F		1.727
Adjusted R^2	0.599	0.599
Number of observations	3,152	3,152

To remove the effect of multicollinearity, we re-estimated the regression without human capital measures of education and foreign population. As shown in Model E in Table 6, removing education and foreign population does not change the model's statistical significance, as the F statistic is still significant at the p=.001 level. Furthermore the R^2 is still significant at the p=.001 and is only slightly smaller than the R^2 in Table 4. This indicates that the two deleted variables are not contributing substantially to the explained variance beyond that of the remaining independent variables. All of the coefficients of the remaining variables, including R&D, continue to have the same signs and similar statistical significance. It should be noted, however, that the R&D coefficient is slightly larger in Model F (0.040) than it was in Model D (0.021) but smaller than that in Model B (0.068).

From this we can conclude that high school education, college education, and foreign population are significant in explaining additional variance in births, but their collinearity with other variables raises questions about the direction and magnitude of their relationships with births. There is no way to resolve this question with the data currently available. However, when all three are removed, the result regarding the effect of R&D remains positive.

Table 6 Regression Results (Collinear Variables Excluded)
Dependent Variable = Births (year t + 1)

Variable	Standardized Coefficients a	
	Model E	Model F
Constant	0 ***	0 ***
Population log	0.136 ***	0.117 ***
Population Change	0.476 ***	0.474 ***
Establishment Density	0.405 ***	0.407 ***
Establishment Size	-0.370 ***	-0.368 ***
Unemployment Rate log	0.074 ***	0.083 ***
Trend	-0.034 **	-0.033 **
R&D Year t - 1		0.040 **
F	872.576 ***	751.235 ***
Change in F		9.950 **
Adjusted R^2	0.596	0.597
Number of Obs.	3,152	3,152

Regressions Using Grants

As a test to see if Grants make any contribution to births, we repeated the regression in Table 6 substituting Grants for R&D. Table 7 shows the results, which cover only the period from 1994, since we have no earlier data on Grants. This is why there are only 1,970 observations.

Grants enters the regression in Model H with a standardized Beta coefficient of 0.058, which is statistically significant at p<0.001. Otherwise, the relative sizes of the control variable coefficients are about the same as in those in Table 6. Also, the R^2 is slightly larger. As an additional test of Grants, we put both grants and R&D into the same model, as shown under Model I. Both Grants and R&D beta coefficients are statistically significant but Grants is much weaker here than in Model H. R&D resumes it previous magnitude and statistical significance as in Model F. The R^2 is only slightly large than that in Model H. And, the F statistic, although still significant, has declined in magnitude. The relative size of the control variables' coefficients remains the same, in spite of the shorter time period and fewer observations. Thus, it appears that because of the large Pearson correlation coefficient between R&D and Grants, the magnitude and significance of the beta coefficients are unreliable. So Model F in Table 6 probably presents the best representation of the relationship of R&D and Births.

Table 7 Regression Results
Dependent Variable = Births t + 1
Added Independent Variables = R&D and Grants

Variable	**Standardized Coefficients**		
	Model G	**Model H**	**Model I**
Constant	0 ***	0 ***	0 ***
Population log	0.155 ***	0.126 ***	0.112 ***
Population Change	0.524 ***	0.521 ***	0.521 ***
Establishment Density	0.439 ***	0.435 ***	0.439 ***
Establishment Size	-0.329 ***	-0.327 ***	-0.324 ***
Unemployment Rate log	0.128 ***	0.137 ***	0.146 ***
Trend	-0.062 ***	-0.061 ***	-0.060 ***
Grant t		0.058 ***	0.042 *
R&D t – 1			0.045 **
F	610.301 ***	528.317 ***	464.704 ***
Change in F		13.359 ***	7.382 **
Adjusted R^2	0.650	0.652	0.653
Number of Observations.	1,970	1,970	1,970

HYPOTHESIS TWO: ECONOMIC GROWTH REGRESSIONS RESULTS

H2 LMA economic growth rates will be positively related to(a) LMA new firm formation rates, and (b) LMA university research and development expenditures, and (c) LMA human capital.

Introduction

We take two approaches to determine the effect of R&D upon economic growth. Given the extensive use of net employment change as a measure of economic growth (Birch, 1979, 1987; Kirchhoff, 1994; Baldwin, 1995), we choose this as our dependent variable. We begin by measuring economic growth by the percent change in LMA employment from year t to year t+1. This growth is then regressed with the independent variables in time t. The results are not satisfactory. The most likely reason is that the time lags between the independent variable and the dependent variables are most likely to be greater than one year and may be different for each independent variable.

To adjust for this, we divided the ten annual observations into two time periods, 1990-1994 and 1995-1999. Then we tested regressions using 1990-94 for the independent variables and 1995-99 for the dependent variable. This “time smear” may reveal additional meaningful results.

Annual Economic Growth Regression

We begin by examining the Pearson correlation matrix of dependent and independent variables for large coefficients that suggest multicollinearity. These are shown in Table 8.

Table 8 Pearson Correlations for
Economic Growth Variables
N = 3,152

	Variable	1	2	3	4	5	6	7	8	9
1	Birth Rate									
2	High School Education	0.180 (0.000)								
3	Unemployment Rate log	0.014 (0.405)	-0.443 (0.000)							
4	Establishment Size	-0.409 (0.000)	0.040 (0.017)	-0.286 (0.000)						
5	Establishment Density	0.501 (0.000)	0.540 (0.000)	-0.419 (0.000)	-0.294 (0.000)					
6	R&D t-1	0.076 (0.000)	0.351 (0.000)	-0.239 (0.000)	0.270 (0.000)	0.088 (0.000)				
7	Employment Change t+1	0.275 (0.000)	0.072 (.0.000	-0.060 (0.000)	-0.087 (0.000)	0.051 (0.003)	0.022 (0.184)			
8	Nat'l GDP Growth Rate	-0.005 (0.771)	0.000 (1.000)	-0.236 (0.000)	0.090 (0.000)	0.086 (0.000)	0.030 (0.072)	0.127 (0.000)		
9	Population log	0.061 (0.000)	0.332 (0.000)	-0.105 (0.000)	0.488 (0.000)	0.052 (0.002)	0.486 (0.000)	-0.031 (0.067)	0.020 (0.227)	
10	Population Growth Rate	0.486 (0.000)	0.120 (0.000)	0.017 (0.322)	0.082 (0.000)	0.032 (0.059	0.133 (0.000)	0.238 (0.000)	0.009 (0.612)	0.196 (0.000)

The highly correlated variables are limited to a few cases. Following earlier patterns, we find that high school education, unemployment rate, establishment density, establishment size are among the correlated independent variables as is birth rate and population growth rate.

However, when we estimate several regression models with change in employment as the dependent variable, we find that they explain very little of the annual variance in employment change rates in LMAs. Table 9 displays the results of two of several regressions, none of which provides any improvement in explained variance. As shown in both models J and K in Table 9, the adjusted R^2 is only 0.130, showing that thirteen percent of the variance is explained.

Table 9 - Regression Results
Dependent Variable = Annual Employment Growth

Variable	Standardized Coefficients	
	Model J	Model K
Constant	0 **	0 **
Birth Rate	0.310 ***	0.310 ***
Establishment Density	-0.187 ***	-0.187 ***
High School Education	0.144 ***	0.144 ***
Establishment Size	0.015	0.015
Nat'l GDP Growth Rate	0.144 ***	0.144 ***
Population log	-0.117 ***	-0.117 ***
Population Growth Rate	0.096 ***	0.096 ***
R&D t		.001
F	73.352***	65.915 ***
Change in F		0.002
Adjusted R2	0.130	0.130
Number of Observations.	3,152	3,152

This result is not surprising since in this regression, we are hypothesizing that economic growth in time t+1 is a function of economic conditions in time t. It is well known that economic growth is a timedelayed function of causal variables and the delay differs from variable to variable. Thus we need to look at a longer period of time in order to test this hypothesis properly.

Longer Term Economic Growth Model

To provide for a longer period of growth, in response to a longer period of economic and social conditions, we chose a five year "time smear," which would allow for the different leads/lags of various explanatory variables. We hypothesize that economic growth in the time period 1995 through 1999 will be function of independent variables in the previous time period, ending in 1994. We have used 1994 data for relatively stable variables describing local conditions, and averages of several years for less stable variables, calculated for the 394 LMAs as described below.

Longer Term Independent Variables and Their Transformations

Average Firm Birth Rate – 1993-94

We calculated this as the average of births in 1993-4 in each LMA. This was the sum of local firm births in the two years divided by two. This was multiplied by 1000 and divided by the average LMA labor force for 1992-4.

Establishment Density - 1994

This was calculated in each LMA by dividing the number of establishments by population divided by 1000 for 1994.

Establishment Size – 1994

Areas dominated by larger establishments are likely to have slower growth, because larger businesses typically have been built to their optimal scale, and have little capacity for expansion. Average establishment size is calculated as total employment in the LMA, divided by total number of establishments in 1994, just prior to our growth period.

High School Education - 1990

Because these data are not available for any more recent year, this was calculated as the number of adults over 24 in 1990 with a high school diploma divided by total population in 1990.

College Degrees - 1990

Again, because no more recent data are available, this was calculated as the number of adults over 24 with a college degree in 1990 divided by total population in 1990.

Population log- 1994

This was taken from the Bureau of the Census county population reports for 1994 and was aggregated into LMAs. The natural logarithm was taken to represent population differences in terms of their proportional differences, rather than absolute numbers.

University R&D – 1993-94

This was calculated as the mean of NSF's R&D expenditure surveys for 1993-94 divided by labor force. The natural logarithm was taken to represent differences in proportional terms.

Population Change – 1993-94

This is the compound rate of growth for population from 1992 through 1994, which is assumed to represent the underlying trend in local growth in the early part of the nineties.

SBIR & STTR Grants 1993-94

This is a dummy variable. It is set equal to one when an LMA has received an SBIR and/or an STTR grant any time during the period of 1993-94. It is zero otherwise.

Longer Term Dependent Variables

Employment Change – 1995-99

We calculated the average compound rate of growth for employment from 1995 through 1999 for each LMA. Take the fourth root of the ratio of 1999 employment divided by 1995 employment, and then subtract 1.0, to construct the average annual rate of change over this period.

Results Hypothesis Two - Long Term Growth

Table 10 below provides the descriptive statistics on these variables.

Table 10 Descriptive Statistics of Variables Used in the Long Term Economic Growth Regressions

N=364

Variable	Minimum	Maximum	Mean	Standard Deviation
Average Births 1992-94	1.952	9.352	3.626	0.903
Establishment Density 1994	11.553	45.105	21.868	3.542
High School Educ.1990	0.459	0.883	0.721	0.080
Establishment Size 1994	8.214	22.945	15.226	2.854
Population 1994 log	11.537	16.538	12.789	0.936
R&D/LF 1993-94 log	0.000	8.235	2.161	2.451
Population Change 1992-4	-0.020	0.062	0.011	0.010
College Degrees 1990	0.069	0.320	0.159	0.050
Employment Change 1995-99	-0.020	0.071	0.020	0.014

Table 11 describes the Pearson correlation coefficients for these variables. Note that 11 of the 36 correlation coefficients have values greater than 0.40. However, in most cases, the correlation coefficients are not above 0.50. Still, we must be wary of the considerable potential for multicollinearity problems.

- - - This area left blank for proper pagination. - - -

Table 11 Pearson Correlation Coefficients on Variables Used in Economic Long Term Growth Regressions

N = 364

	Variable	1	2	3	4	5	6	7	8
1	Employment Change								
2	Establishment Size 1994	0.004 (0.938)							
3	Average Births 1993-94	0.436 (0.000)	-0.438 (0.000)						
4	Establishment Density 1994	0.183 (0.000)	-0.335 (0.000)	0.535 (0.000)					
5	High School Education in 1990	0.358 (0.000)	0.018 (0.718)	0.161 (0.001)	0.536 (0.000)				
6	College Degrees 1990	0.432 (0.000)	0.181 (0.,000)	0.245 (0.000)	0.385 (0.000)	0.701 (0.000)			
7	R&D /LF log 1993-94	0.232 (0.000)	0.251 (0.000)	0.061 (0.224)	0.074 (0.142)	0.350 (0.000)	0.701 (0.000)		
8	Population Change 1992-94	0.548 (0.000)	0.011 (0.827)	0.527 (0.000)	0.044 (0.388)	0.116 (0.021)	0.228 (0.000)	0.120 (0.017)	
9	Population log 1994	0.307 (0.000)	0.465 (0.000)	0.040 (0.428)	0.033 (0.509)	0.333 (0.000)	0.609 (0.000)	0.483 (0.000)	0.100 (0.048)

Regression Results – Employment Change

Table 12 below shows the results of estimating regression for four variations on our basic growth model. Model L describes the basic model, with human capital represented only by adult high school attainment. The explained variance for this regression is satisfactory with an R^2 of .438. Note the large standardized coefficients for average births and average population change and high school education. These three variables together explain a great deal of the variance in employment change in all of the five models we estimate here.

The Pearson correlations show a large coefficient between high school education and average firm birth rates, suggesting that multicollinearity may exist with these two variables so that their coefficients may be unreliable. To test this, we substituted college degrees for high school education in Model O, and we find that college degree attainment is statistically significant, but that the explained variance (R^2) falls, relative to that for the models with high school attainment. Firm births has a smaller beta coefficient in Model O, suggesting that its effect on economic growth may be less than indicated by the size of its beta coefficient in the other four models in Table 12. Nonetheless, births still has a large and significant coefficient in Model O.

University R&D is added to the basic Model L to form Model M. The University R&D adds nothing to the regression. Its coefficient is small and non-significant and the Model M statistics are very similar to Model L. The adjusted R^2 actually declines slightly due to the addition of a variable without any increase in the explained variance. Thus, we conclude that University R&D expenditures do not contribute significantly to explaining the variance in the rates of change in local employment. Since R&D does not have a large Pearson correlation coefficient with any of the variables in the basic Model K, this lack of significance cannot be due to multicollinearity.

Model N shows the same regression variables without university R&D but with the SBIR and STTR Grants added. Again, as with university R&D, Grants contributes no additional explained variance and is not significant.

Table 12 Long Term Regression Results
Dependent Variable = Employment Change 1995-99
N = 394

Variables	**Standardized Beta Coefficients**				
	Model L	Model M	Model N	Model O	Model P
Constant	0 ***	0 ***	0 ***	0 ***	0 ***
Average Births – 1993-94	0.262***	0.262***	0.262***	0.186***	0.246***
Population Change – 1992-94	0.365***	0.365***	0.367***	0.385***	0.373***
High School Education – 1990	0.290***	0.290***	0.294***		0.288***
College Degrees – 1990				0.239***	
Population Log – 1994	0.168***	0.168***	0.176***	0.116*	0.181***
Establishment Density – 1994	-0.134*	-0.134*	-0.134*	-0.029	-0.134*
Establishment Size 1994					-0.027
R&D/LF log – 1993-94		0.000			
SBIR & STTR Grants – 1993-94			-0.019		
F	62.24	51.73	51.79	48.04	51.81
Adjusted R^2	0.438	0.436	0.437	0.418	0.437

Model O incorporates College Education in place of High School Education. In this regression, the coefficient for Population Log becomes much smaller and less significant. This may be due to the high degree of multicollinearity between Population Log and College Education, as suggested by their Pearson correlation coefficient of 0.609. Multicollinearity may cause the importance of one or both variables' coefficients to be unreliable. Moreover, the R^2 in Model O is smaller (0.418) than in the other four models indicating that less variance is explained with this Model. Model P shows the same variables as in Model L but with establishment size added. This reduces the R^2 slightly and is not statistically significant. It adds nothing to the explained variance.

An important aspect of all these models is that both average firm births and human capital (either high school or college education) have relatively large standardized coefficients at a high degree of significance in all five regressions.

DISCUSSION AND CONCLUSIONS

We begin our discussion with the results of the highly successful tests of hypothesis one, examining the relationship of R&D expenditures with firm birth rates. Next, we discuss the outcome of our much more difficult analysis of economic growth rates, which is summarized in hypothesis two. We conclude with an overview of what we have found and the public policy implications that emerge from our research.

Hypothesis One Discussion

H1 LMA new firm formation rates will be positively related to (a) LMA university research and development expenditures and (b) LMA human capital.

R&D Expenditures

The regression model I displayed in Table 7 and Model F in Table 6 demonstrate that LMA university R&D expenditures have a statistically significant local effect on firm births, and that the strongest relationship is with a two-year lag between the fiscal year of the expenditures and the calendar year of the firm births. For other explanatory variables it was assumed that conditions in time t, or local rates of change during the period ending in time t, affect the rate of new firm births measured in time t+1, as the year in which they first have employees. Moreover, SBIR and STTR grants (in time t) also have a statistically significant effect on firm birth rates, though not as strong. These results support part (a) of hypothesis one.

These results also suggest that Jaffe (1986) type "spillover" effects are at work in and around universities. And, where university R&D is larger, this effect is greater, just as one would expect in an R&D intensive industries. New firms form around the local "university" research activity centers much as Audretsch and Feldman (1996) found in for certain local industry clusters. These university activity centers spawn new firms in much the same way as local industry clusters.

Human Capital

However, this regression does not show human capital as a significant variable in the model. Part (b) of hypothesis one finds no support here except in Models C and D with College Education. High School Degrees is statistically significant in Models A and B but with an opposite sign (negative) that can not be explained except as a result of multicollinearity. It is probable that multicollinearity of human capital variables (high school and college education) causes these variables to be non-significant. The large Pearson correlation between the human capital measures and R&D suggests that human capital is related to R&D and to births. But, multicollinearity forces us to accept one or the other variable into the model. We do this by omitting both from model E and F in Table 8. Here we experience very little loss of explained variance in Model E even though both human capital variables are omitted. Then, we add R&D to Model E to form Model F where R&D's addition shows an increase in explained variance. This confirms R&D's effect on births but questions human capital's contribution.

Moreover, the proportion of adults with college degrees has a large Pearson correlation with population. This is not surprising since it is well known that college educated adults tend to congregate in the larger metropolitan areas. High school education has large Pearson correlation coefficients with establishment density and unemployment rate. And, high school education appears with a small and negative coefficient in Models A and B in Table 4. This is thought to be an erroneous coefficient magnitude and sign due to multicollinearity with establishment density and unemployment. College education has a small and positive coefficient, when substituted for high school education in Models C and D. Introducing college education into the regression causes the R^2 to decline from 0.614 in Model A and 0.617 in Model B to 0.599 in Models C and D. This suggests that multicollinearity may have disrupted the regression results. Even high school education is removed in our final regression because of its high correlation and likely multicollinearity with establishment density and unemployment.

Thus, we find no definitive evidence that human capital is related to economic growth and admit that multicollinearity may be at fault for failing to find such evidence.

Other Exogenous Variables

The finding that population has a significant and positive relationship to the rate of new firm births in LMAs shows that the agglomeration effects of larger concentrations of mobile population in the larger LMAs tends to promote new firm births. Apparently the larger economic areas have more specialized markets and skills available for new businesses, or better functioning structures to promote spillover of the new knowledge being discovered within them, or more competitive attitudes about business and entrepreneurship, any of which would contribute to higher rates of innovation and new business formation.

The rate of change in population of LMAs is also statistically significant and positively related to firm birth rates. At the most basic level, since people form businesses, the more people in the area, the more businesses will be formed, so the faster the human population grows, the faster the rates of new business creation. However, this suggests that the possible negative effects of crowding – higher costs due to shortage of land and infrastructure – are offset by the positive effects of a greater supply of both consumers and workers. In addition, the recent growth rate of an area is probably the best indicator of its relative desirability as a place to locate both homes and businesses. This result is also is consistent with our expectations.

Establishment density, measured as LMA total establishments divided by the population, is related to firm births. Since we chose establishment density as a surrogate for the extent of competition, we expected that this would be negatively related under the assumption that greater competition will discourage firm births. Apparently, the spillover effects from existing establishments have greater influence on new firm births than competitive effects. This result may be explained by noting that entrepreneurs founding new firms based on innovations often do not believe their innovations have competition since the uniqueness of their innovations will take market share away from existing suppliers.

Establishment size, measured as total employment divided by total establishments, is significant and negative as we expected. As the average size of establishments increases, we expect to find more and more large organizations in the LMA and lower innovation rates and firm birth rates. The lower rate of innovation in larger organizations is confirmed by Gellman (1976), the Futures Group (1984), and Chakrabarti (1991). And a lower innovation rate reduces the spillover effects that underlie firm births.

The unemployment rate is significant and negative as expected. We selected unemployment as an indicator of overall economic activity in the LMAs. Although it is often said that unemployment stimulates entrepreneurship, Reynolds et al. (1995) found that there was a very small percentage of entrepreneurs who were stimulated to start a firm because of unemployment. Our results agree with this finding and also demonstrate that poor economic conditions cause a decline in firm birth activity. This agrees with Kirchhoff's (1994) observations on the U.S. economy as a whole.

Hypothesis Two Discussion

> *H2 LMA economic growth rates will be positively related to (a) LMA new firm formation rates, and (b) LMA university research and development expenditures, and (c) LMA human capital.*

This hypothesis presents many more problems that hypothesis one. We select net new job creation as the measure of economic growth because it has been the most widely used indicator of growth in the literature of small business and economic growth (Birch, 1979, 1987; Kirchhoff, 1994; Baldwin, 1995; Wennekers and Thurick, 1999). Because the leads and lags of environmental variables vary considerably, it is not possible to find an annual relationship between net new job creation in the year following that of the independent variables. Therefore we focused on investigating the longer term relationship between the average local growth rate in 1995 through 1999 and the values of the independent variables prior to that period All variables were transformed to accommodate this change to a longer term relationship.

Firm Birth Rates

Regressions on this revised database show than the exogenous factors of population, population growth rate, human capital – both high school and college when examined separately – and establishment density had consistent statistical significance in all regressions. And, the direction of their effects is as we expected, all positive. But, most important, average firm births has a statistically significant and large, positive standardized beta coefficient. Thus, the first part of our hypothesis is confirmed, i.e. firm birth rates have a positive impact upon local economic growth.

University R&D Expenditures

However, university R&D expenditures are not significant in any of the models of employment growth that also include firm birth rates as an independent variable. Nor are SBIR and STTR grants statistically significant. Thus, we can conclude that university R&D has no detectable effect on economic growth rates, and reject the part (b) of Hypothesis two.

However, since R&D does have an effect on firm births as shown in the test of hypothesis one, and firm births have a substantial effect on local economic growth, we can conclude that these results suggest that university R&D expenditures influence local economic growth through the birth of new firms.

Human Capital

Human capital measured as high school education and college degrees is a significant variable in all five models in Table 12. However, the large correlation coefficient (0.701) between these two measures means that either one but not both can be included in the model. Since high school education has the larger standardized beta coefficient, it is the logical choice. Thus, hypothesis two (c), that LMA economic growth rate will be related positively to human capital, is supported by this regression analysis.

Other Exogenous Variables

Three of the environmental variables appear in the regression models in the same way that they appear in the regressions on hypothesis one. Population, population change and establishment density are all present. Population and population change are related positively as expected.

Some what surprisingly, establishment density is now related negatively to economic growth. This seems to corroborate are original thinking that this variable represents the degree of competition in the LMA. This suggests that greater competition results in lower growth in employment. However, establishment density is barely significant ($0.01<p< 0.05$) in four of the five regressions. Establishment density also has large Pearson correlations with average births, high school education and population. In model O, high school births is replaced by college degrees and establishment density is not significant. This leads us to conclude that both the sign and magnitude of establishment density's beta coefficient may be an unreliable indicator of its relationship with economic growth because of multicollinearity.

Conclusions and Policy Implications

University expenditures on research and development promote higher new firm birth rates. The phenomenon is identical to that described by other researchers as "spillover" effects. Just like business firms, research universities form local innovative activity centers, from which both knowledge spillovers and growth in specialized markets generate higher rates of new firm formation in one or more industries.

The glue that holds these clusters together is the effort universities are putting into mechanisms to promote commercialization of the inventions that emerge from their laboratories. Such mechanisms include campus incubators where new firms can obtain low overhead space that includes low cost shared services; professorial sabbaticals to develop inventions into commercial products or services; development of science parks; assistance with patent applications; and even university-created venture capital funds poised to provide the early, start- up stage capital that is so scarce in the private sector.

Beyond this, our research shows that the effect of R&D expenditures on firm births endures annually for at least five years after the R&D spending occurs. This effect is not only due to the growth of the newly formed entrepreneurial firms based on inventions derived from university R&D, but also to the secondary effects of their employment growth. Economists differentiate between primary and secondary growth effects. Primary effects are the direct increase in new jobs due to employment in the newly formed R&D based firms. Secondary effects are the employment increases that occur due to birth and growth of new firms that support the primary R&D-based new firms, and to serve the families of workers that settle nearby to benefit from those employment opportunities. These secondary new firms may be suppliers to the primary firms. Or the secondary births may be retailers who supply the needs of the newly employed workers in the R&D based firms. Since our birth measurement includes all new firms, regardless of industry, we have captured both primary and secondary effects in our birth measure.

We are not the first to provide evidence of this effect. After studying regional growth in large metropolitan areas, small metropolitan areas and rural areas annually for five years, Birch, Haggerty, and Parsons (2000) report that the factors most important to economic growth are in order of importance:

1. Universities
2. Skilled labor pool
3. Airports
4. A nice place to live.

"Universities are the feedstock for most Gazelles" (i.e. high growth, new firms). University labs are frequently the source of the entrepreneur" (Birch, et al., 2000, p. 9). Although Birch et al. use less rigorous geographical area designations than we use herein, the database they use is as expansive in coverage as Census' LEEM and they analyzed the same time period. [7] Our results using LMA data are consistent with Birch et al.'s conclusions.

Beyond their direct involvement in such firm birth mechanisms, universities also feed highly trained personnel into these newly formed firms. Many graduate and under-graduate degrees are awarded to students who, as part of their formal studies, worked for faculty engaged in the invention process. "Our calculations suggest that a high proportion of growing firms hire a workforce with above average skill levels" (Birch, et al., 2000, p. 9). The large Pearson correlation coefficient between college educated persons and R&D expenditures that is so evident in this report suggests that research universities graduate and attract college graduates to their local areas to provide the educated work force needed to support the clustering of new firms. At the same time, these new firms are able to attract college graduates with promotion opportunities accompanying rapid growth.

Research universities are significant contributors to the national economy. Their research efforts create new firms that grow and create new jobs. Public policy needs to look at this role universities play in new firm creation. The important question is whether or not increased spending on University R&D would increase the birth rate of new firms and therefore, economic growth.

By NSF's measures, in the year 2000, 58 percent of universities' expenditures on R&D are funded by the federal government. Thirty-four percent are funded by states, foundations, or the universities themselves, with only eight percent funded by private businesses. However, the 1990 through 1997 compound rate of growth for university R&D expenditures was 3.2 percent per year compared to the twenty year average of 4.9 percent (constant dollars). During this time, the federal government's share of R&D declined while that of industry increased. Our results suggest that increasing the federal investment in R&D will not have a demonstrable effect until one year after the R&D is expended. And then, the full effect will not be realized until five or six years later. On the other hand, the full effect of the gradual decline in federal R&D share as described above will not show in reduced firm births for five or six years. This lagged effect makes relating economic vitality to R&D expenditures difficult and makes it more difficult to convince policy makers that university R&D is a wise investment for the future of the economy.

A recent federal government report (Office of Technology Policy, 2000, p1-1) also emphasizes the need for investments to increase the research capacity of areas, to accelerate knowledge creation, which will increase growth of the local economies directly, as well as build their sustainable competitive advantages for long-term growth. They note that all of the areas of technology-based economic development in the U.S. have strong concentrations of research, both university and private-based. The Office of Technology Policy report provides detailed state-level data for further analysis of these relationships, but too often very little correlation exists between state boundaries and economic areas. Thus, it is very difficult to investigate these relationships in more detail at the state level. Never the less, it is clear that the wide disparity in university research funding among the states is closely related to their differences in economic growth rates. Funding of university research staff and facilities, and support for a better educated labor force, are both valuable instruments for triggering higher local growth rates, and greater local and national competitiveness. The research reported here adds empirical evidence to support the validity of the Office of Technology Policy's assertion.

There are LMAs that have no university among the 600 universities covered by the NSF survey of research expenditures. These LMAs will not experience the economic wealth experienced by those LMAs that have one or more research universities. State economic development specialists in the disadvantaged LMAs would be wise to consider using economic development funds to establish and promote at least one public university in such LMAs to become a research university so that real economic growth can be accelerated. This will take time but in the long run the economic growth will benefit a larger number of citizens.

[7] This information is taken from personal conversations between Bruce Kirchhoff and David Birch, June 2002.

And, more research is required. We hope to exhaust this database with various approaches, multi-stage regression, multiple equation models, and other sophisticated research tools shall be applied in the coming years.

REFERENCES

Acs, Zoltan J., David B. Audretsch, and Maryann P. Feldman. 1994. R&D Spillovers and Recipient Firm Size. *The Review of Economics and Statistics*. 76(2): 336-340.

Almus, M, and E. A. Nerlinger, 1999, Growth of New Technology Based Firms: Which Factors Matter?, *Small Business Economics*, Volume 13 (2), pp 141-154

Anselin,L., A. Varga, and Zoltan J. Acs, 1997. Local Geographic Spillovers between University Research and High Technology Innovation, *Journal of Urban Economics*, 42, 422-448.

Armington, Catherine and Zoltan J. Acs. 2002. The Determinants of Regional Variation in New Firm Formation. *Regional Studies*, Volume 36, pp 33-45.

Audretsch, David B. 1991. New-Firm Survival and the Technological Regime. *The Review of Economics and Statistics*. 73(3): 441-450.

Audretsch, David B. 1995. Innovation, Growth and Survival. *International Journal of Industrial Organization*. 13: 441-45

Audretsch, David B. and Maryann P. Feldman. 1996. R&D Spillovers and the Geography of Innovation and Production. *The American Economic Review*. 83(3): 630-640.

Baldwin, John R., 1995, *The Dynamics of Industrial Competition*, Cambridge University Press.

Birch, D. 1987. *Job Creation in America: How our smallest companies put the most people to work.* New York, The Free Press.

Birch, David L., 1979, *The Job Generation Process*, Unpublished report prepared by the Massachusetts Institute of Technology Program on Neighborhood and Regional Change for the Economic Development Administration, U.S. Department of Commerce, Washington, D.C.

Birch, David, Anne Haggerty, and William Parsons, 2000, *Entrepreneurial Hot Spots: The Best Places in America to Start and Grow a Company.* Cambridge, MA: Cognetics, Inc.

Chakrabarti, Alok K. 1991. Industry Characteristics Influencing the Technical Output: A Case of Small and Medium-Sized Firms in the U.S. *R&D Management*. 21(2): 139-152.

Futures Group, The. 1984. *Characterization of Innovations Introduced on the U.S. Market in 1982.* A report prepared for: Washington, D.C.: U.S. Small Business Administration, Office of Advocacy.

Gellman Research Associates, 1976, *Indicators of International Trends in Technological Innovation*. A report prepared for: Washington, D.C.: U.S. Small Business Administration, Office of Advocacy.

Glaeser, Edward L., 2000, The New Economics of Urban and Regional Growth, in *The Oxford Handbook of Economic Geography*, Gordon L. Clark, Maryann P. Feldman and Meric S. Gertler eds.: The Oxford University Press, Oxford, England; p. 83-98.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. 1992. Growth in Cities. *Journal of Political Economy*. 100(6): 1126-1152.

Henderson, Vernon, Ari Kuncoro, and Matt Turner. 1995. Industrial Development in Cities. *Journal of Political Economy*. 103(5): 1067-1090.

Jaffe, Adam B. 1989. Real Effects of Academic Research. *The American Economic Review*. 79(5): 957-970.

Jaffe, Adam B. 1986. Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value. *The American Economic Review*. 76(5): 984-1001.

Kirchhoff, Bruce A. 1994. *Entrepreneurship and Dynamic Capitalism: The Economics of Business Firm Formation and Growth*. Westport, CT: Praeger.

National Science Board, 2002, *Science Indicators and Engineering Indicators – 2002*, Arlington, VA: National Science Foundation (NSB-02-1).

Office of Technology Policy, 2000, *The Dynamics of Technology-based Economic Development: State Science and Technology Indicators*. Technology Administration, U.S. Department of Commerce.

Reynolds, Paul D., Brenda Miller, and Wilbur R. Maki. 1995. Explaining Regional Variation in Business Births and Deaths: U.S. 1976-88. *Small Business Economics*. 7: 389-407.

Scherer, F. M. 1984. *Innovation and Growth: Schumpeterian Perspectives*. Cambridge, MA: The MIT Press.

Schumpeter, Joseph A., 1934 (1983), *The Theory of Economic Development*, New Brunswick, NJ: Transaction Publishers .

Storey, David. J., 1994, *Understanding the Small Business Sector*, London: Routledge.

Tesfaye, Besrat, (1997) Patterns of Formation and Development of High-Technology Entrepreneurs, Dylan Jones-Evans and Magnus Klofsten, eds., *Technology, Innovation and Enterprise*, St Martin's Press, New York, NY, p. 61-106.

Tolbert, Charles M. and Molly Sizer (1996) *U. S. Commuting Zones and Labor Market Areas: a 1990 Update*, Washington, D.C.: Economic Research Service, Rural Economy Division, U.S. Department of Agriculture.

Wennekers, S, R. Thurik, 1999, "Linking Entrepreneurship and Economic Growth," Small Business Economics, Volume 13 (1), pp 27-55