Using Census BITS To Explore Entrepreneurship, Geography, and Economic Growth

by
Zoltan C. Acs and Catherine Armington
Acs & Associates
Ruxton, Maryland 21204

for

under contract number SBA-HQ-03-M0534

Release Date: February 2005

SMALL BUSINESS
RESEARCH SUMMARY

SBA Office of Advocacy

No. 248
February 2005

Using Census BITS To Explore Entrepreneurship, Geography, and Economic Growth

by Zoltan C. Acs and Catherine Armington;
Acs & Associates [45] pages. Ruxton, Maryland 21204
under contract no. SBA-HQ-03-M0534

This paper discusses three related research projects that used the Business Information Tracking Series (BITS). This set of data was developed by the U.S. Bureau of the Census with support from the U.S. Small Business Administration Office of Advocacy. The purpose of the report is to demonstrate the importance and usability of large micro databases such as the BITS that cover almost all U.S. private sector businesses, and track their employment and firm ownership annually from 1988 through 2001. It represents all industries, distinguishes between establishments and firms, identifies start-ups of both new firms and new secondary locations of old firms, and specifies the location, industry, and changing employment of each establishment through time.

Overall Findings

Empirical results using BITS data in three studies bore out the authors' contention that models of economic growth would be substantially improved by greater incorporation of theories of entrepreneurial activity. Variation in firm formation rates (defined as the number of net new firm divided by 1,000 labor force in a particular year) is substantially explained by regional differences in industry specialization, human capital, and local population and income growth rates.

Highlights

• The underlying theoretical framework for the empirical analysis includes three components: spatially constrained externalities, increasing returns in the production of goods, and decreasing returns in the production of knowledge.

• Using annual data on firm formation for 384 labor market areas, in six industry sectors between 1991-1998, the researchers found considerable variation in the new firm formation rate across regions, but little variation over time.

• The results suggest that the regional differences in service firm formation rates depend to a large degree on the educational requirements and the market served by the newly formed firms.

• From 1996-1999, new firm formation rate increases are associated with higher employment rate growth. This relationship holds for all periods examined and for all sectors of the economy, except for manufacturing.

• An increase in the number of high school graduates tended to increase overall growth rates for the period of 1996-1999. The research also find evidence that raising the overall level of education (high school graduation) has a greater impact on economic growth than raising the proportion of the adult population with college degrees, if new firm formation rates are held constant.

• Firm formation rates tend to be higher in regions with higher business specialization rates. In addition, it finds that greater physical density, defined as the number of establishments per square mile, of businesses in a region is also associated with less, not more, growth.

• Other longitudinal business microdata bases are available, but none is as useful as the BITS for tracking firms, and their establishments and employment dynamics. No other can detect new firm formations properly, since the United States. lacks a national system for public registration of businesses. In addition, more and more businesses are not bothering with Yellow Page subscriptions or commercial credit applications.

• The primary limitation of the BITS data is that they are only available from 1988, with longitudinal linkage from 1989. As with all large business databases, the BITS inevitably has some imperfections in its tracking of establishments that are changing their ownership or legal form over time, and of firms that are creating, divesting, or closing secondary locations. Despite these weaknesses, the BITS provide far more accurate and comprehensive data on U.S. private sector businesses than any other source.

Data Sources and Methodology

The current Business Information Tracking Series data for 1990 to 1999 have been used to estimate regression models to test hypotheses from recent literature that are related to variables such as entrepreneurial activities, industrial specialization, human capital, localities, and economic growth. The final report was peer reviewed consistent with the Office of Advocacy's data quality guidelines. More information on this process can be obtained by contacting the director of economic research at *advocacy@sba.gov* or (202) 205-6533.

Ordering Information

The full text of this report and summaries of other studies performed under contract to the U.S. Small Business Administration's Office of Advocacy are available at *www.sba.gov/advo/research.* Copies are also available from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
(703) 487-4650, (703) 487-4639 (TDD)
www.ntis.gov
NTIS order number: PB2005-102189
Pricing information: Pending

To receive email notices of new Advocacy research, press releases, regulatory communications, and publications, including the latest issue of *The Small Business Advocate* newsletter, visit *http://web.sba.gov/list* and subscribe to the appropriate Listserv.

EXECUTIVE SUMMARY

- This paper discusses three related research projects that use the Business Information Tracking Series (BITS) that was developed by the U.S. Bureau of the Census with support from the U.S. Small Business Administration Office of Advocacy. The BITS data cover almost all U.S. private sector businesses, and track their employment and firm ownership annually from 1988 through 2001.

- These microdata facilitate research on the dynamics of American business, especially on their patterns of new firm formation and closure, employment changes and mergers. In order to test hypotheses about how and why regions differ in their rates of firm formation and employment growth one needs a database such as BITS, which represents all industries, distinguishes between establishments and firms, identifies start-ups of both new firms and new secondary locations of old firms, and specifies the location, industry, and changing employment of each establishment through time.

- We examine regional variation in entrepreneurial activity in light of theoretical developments in economic geography and new growth theory. The ability to transform new knowledge into economic knowledge requires a set of skills, aptitudes, insights and circumstances that is neither uniformly nor widely distributed in the population. Thus a closer connection between endogenous growth models and models of entrepreneurship seems necessary. This study used a theoretical framework that has three cornerstones: spatially constrained externalities, increasing returns in the production of goods, and decreasing returns in the production of knowledge.

- Using annual data on firm formation for 384 labor market areas, in six industry sectors between 1991–1998, we find considerable variation in the new firm formation rate across regions, but little variation over time. Variation in firm formation rates is substantially explained by regional differences in industry specialization, human capital, and local population and income growth rates.

- To test the hypotheses that the new firm formation rates are positively related to the level of human capital in the service sector we estimate a model where the dependent variable is the annual service firm formation rates in 1990–1998, or subsectors of service firm formation rates for 1996-98. The results suggest that the regional differences in service firm formation rates depend to a large degree on the educational requirements and the market served by the newly formed firms. The local levels of educational attainment impact primarily the firm formation rates of the types of firms that are normally founded by better-educated entrepreneurs, and do not affect startup rates for those normally founded by individuals with less than a college degree.

- We use BITS data for 1990 to 1999 to estimate a regression model of the regional variation in the employment growth rates to test the hypotheses from recent literature that higher employment growth rates should be positively associated with increased *entrepreneurial activity*, with *higher levels of human capital*, and with more *industry diversity*.

- We find for the more recent period from 1996-99 that if the new firm formation rate increases by one standard deviation, from its mean of 3.5 per thousand-labor force to 4.4 per thousand, the employment growth rate will increase by one-half standard deviation, from its mean of 2.1% to 2.9%. This relationship holds for all periods examined and for all sectors of the economy, except for manufacturing, where new plants may be more important than new

firms, and differences in firm formation rates were not significant in explaining employment growth.

- We find that the importance of local differences in educational attainment increased considerably during the decade. By 1996-99 an increase of one standard deviation in the high school graduation share, from 72% to 80%, accounted for an increase of 0.5% in growth rates. We also find evidence that raising the overall level of education (high school graduation) has a greater impact on economic growth than raising the proportion of the adult population with college degrees, if new firm formation rates are held constant.

- We find that where the business specialization rate is higher by one standard deviation, say, 2.6 establishments per 1, 000 population rather than the mean of 2.2, the employment growth rate is 0.6% lower. However, firm formation rates tend to be higher in regions with higher business specialization rates. We also find that greater physical density of businesses in a region is also associated with less, not more, growth.

- Overall, our results from the three studies support the contention that models of economic growth would be substantially improved by greater incorporation of theories of entrepreneurial activity.

- Other longitudinal business microdata bases are available, but none is as useful as the BITS for tracking firms, and their establishments and employment dynamics. No other can detect new firm formations properly, because the U.S. does not have a national system for public registration of businesses, and increasing numbers of businesses do not bother with Yellow Page subscriptions or commercial credit applications.

- The primary limitation of the BITS data is that they are only available from 1988, with longitudinal linkage from 1989. As with all large business databases, the BITS inevitably has some imperfections in its tracking of establishments that are changing their ownership or legal form over time, and of firms that are creating, divesting, or closing secondary locations. Despite these weaknesses, the BITS provide far more accurate and comprehensive data on U.S. private sector businesses than any other source.

CONTENTS

I. INTRODUCTION

The Business Information Tracking Series (BITS) comprises skeletal data on almost all U.S. private sector businesses that paid any employees during the period from 1989 through 2001. These data have been linked together over time so that there are annual observations on each business establishment, including information on the entire firm (or enterprise) to which it belongs each year. Each establishment is a business location at which goods or services are produced, and these establishments are carefully tracked over time, even as they change legal form, ownership, primary industry, or location. Potentially new businesses are identified from new entries in the Master Business List of the Internal Revenue Service, and Census Bureau programs then use additional administrative data, surveys, economic censuses, and estimates to collect more detailed information on each business and to update this annually.

These microdata facilitate research on the dynamics of American businesses—especially on their patterns of formation, employment change, mergers and acquisitions, and survival. These patterns can be analyzed for establishments in different industry sectors and regions, and for firms of various sizes and other characteristics. They constitute the only comprehensive source of longitudinal microdata that includes the characteristics of the firm to which each establishment belongs.

Acs and Armington have used these BITS data extensively, most recently for a series of research projects investigating regional differences in new firm formation rates and employment growth rates, using those rates as indicators of differences in entrepreneurship levels and local economic growth.[1] Our interest was in how the growth rate differences were associated with the entrepreneurship rate differences and with other socio-economic characteristics of these regions. This type of research is possible only through use of the BITS data, using special tabulations of the microdata to construct an appropriate database covering employment, establishments, and new firm formations by industry and region (394 Labor Market Areas in this case).

In this paper we will first describe the BITS database and review the major research projects that it has been used for. Then we will discuss our use of it for investigating entrepreneurship, geography and economic growth. This is followed by summaries of the empirical results of each of the three sections of this research. We then examine the advantages of the BITS database over other U.S. and foreign data for this type of research. Finally we discuss the current limitations of the BITS data, especially those resulting from its confidential nature.

II. THE BUSINESS INFORMATION TRACKING SYSTEM (BITS)

1) Brief description of the database

The current Business Information Tracking Series (BITS) file facilitates tracking employment, payroll, and firm affiliation and (employment) size for the more than 13 million establishments that existed at some time during 1989 through 2001. This database was constructed from the Census Bureau's Statistics of U.S. Business (SUSB) files, which were developed from the microdata underlying the aggregate data published in Census' County Business Patterns. These annual data describing establishments were linked together using the SUSB Longitudinal Pointer File, which facilitates tracking establishments over time, even when they change ownership and identification numbers. The SUSB data beginning in 1988, their Longitudinal Pointer File, and the BITS files were constructed by Census with substantial support from the Office of Advocacy of the U.S. Small Business Administration.[2]

The basic unit of the BITS data is a business establishment (location or plant). 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. The recorded start year is the year that establishment entered the Census register, which would normally be the year it first hired any paid employees. Additional data for each establishment and year identify the firm (or enterprise) to which the establishment belongs, and the total employment of that firm.

As with most microdata at the Census Bureau, the BITS data are confidential, so the microdata can only be used at the Census Bureau by Census employees, or by approved outside researchers working at one of the Census Centers for Economic Research. These microdata are referred to there as LEEM data (for Longitudinal Establishment and Enterprise Microdata) files. However, many tabulations of these data have been prepared for use by the SBA for other research projects, and these aggregated data are available for further research use.

2) Longitudinal microdata required to evaluate these theories

In order to test hypotheses about how and why regions differ in their rates of formation of new firms and growth of employment, one needs a database representing all industry sectors, that distinguishes business establishments from firms, identifies startups of new firms, and specifies the location and changing employment of each establishment through time. The studies reported here depend crucially on use of the BITS database that the Bureau of the Census has constructed for the Office of Advocacy of the U.S. Small Business Administration for study of entry, survival, and growth in different types of businesses.

This BITS database is a unique by-product of the complex register that Census maintains with information on all businesses in the United States. This Standard Statistical Establishment List, or SSEL, is updated continuously with data from many other sources, but its underlying coverage is based on new business names and addresses from the Master Business File of the Internal Revenue Service. Therefore, every business in the United States that files any tax return is covered by the SSEL, and IRS data from quarterly payroll tax filings (including employment only for the March 12 payroll period) are used to provide comprehensive coverage of all U.S. employment. However, some of the employment numbers are estimated from the payroll numbers, which provide good estimates for establishments that are themselves single-location firms (tax-filing units), but are less reliable for the tax-filing units that represent multiple establishments owned by a single firm.

The data in the SSEL on the individual locations of multi-unit businesses are therefore somewhat less consistent and comprehensive than those for independent, or single-location, firms. All large (over 250 employees reported) multi-unit firms are surveyed annually (Company Organization Survey, except in Economic Census years) to determine the location, industry, and employment (and predecessor or successor owners) of the individual establishments that are or

were owned or controlled by each. A sample of the smaller multi-unit firms is also covered each year, on a rotating basis, so that all but the smallest (with less than 10 employees) are surveyed at least once between each Economic Census, in addition to their coverage in the quinquenial Economic Census (in each year ending with 2 or 7). The resulting lags in the reporting of the formation, closure, or change in ownership of some locations of smaller multi-unit firms (as well as formerly single-unit firms that have become multi-unit) causes false jumps in some of their establishment employment data, as the temporarily aggregated employment of the covered locations is subsequently correctly distributed to the updated list of actual locations.[3]

Census' annual County Business Patterns (CBP) publication provides aggregations of data on establishments selected from the SSEL and extensively edited at both the establishment level (relative to the previous year's data) and the aggregate level. This CBP subset of the SSEL population represents all active (with positive annual payroll) private sector establishments except those in agricultural production, railroads, large pension, health and welfare funds, and private households. The numbers of establishments, employment, and payroll are classified by state, industry, and employment-size class and then processed to avoid disclosure of confidential data, and published annually as CBP.

The microdata behind the CBP provide the starting point for each annual SUSB file. These are further processed to calculate Metropolitan Statistical Area (MSA) codes, and to improve industry code reporting using the industry codes from the subsequent year SSEL whenever they are more precise. Then firm-level data are constructed by aggregating data from all establishments belonging to each enterprise (industry-wide and country-wide), and these firm-level data are attached to each of the component establishment records. These firm-level data include employment, payroll, and receipts (unedited), with the primary state, primary industrydivision, and primary (3-digit SIC) industry within the primary division, based on the largest share of annual payroll. These SUSB data are tabulated and processed for disclosure for a number of standardized tables by firm-size for the Office of Advocacy of the SBA, and for the aggregate SUSB (public) database of the Census Bureau.

A Longitudinal Pointer File is then constructed to link each year's establishment record to the prior year's record for the same establishment, allowing for a change in identity or ownership of continuing establishments. The CFN is the basic Census identification number, which is assigned to each new establishment, and is generally retained consistently over time. However, a change in ownership or legal form, or a change in status between single-unit and multi-unit, will cause a change in CFN. A complex system of computerized matching of records for establishments that might have changed CFN's is used to identify continuing establishments in the SUSB and to update the Longitudinal Pointer File each year. This system examines a wide variety of information, including Permanent Plant Numbers (PPN's), Employer Identification Numbers (EIN's), and statistical matching of records for single units, based on such attributes as name, address, zip code, and industry codes. Matches are sought both between years, and within years (mid-year reorganizations).4 The records that remain unmatched are assumed to represent new establishment formations or closures of existing establishments.

The BITS files are constructed by merging annual SUSB files using the Longitudinal Pointer Files to create a single longitudinal record for each establishment that appears in any of the annual files. Where there has been a mid-year reorganization the data from the two records representing the same establishment are combined for that year, and both CFN's are retained in the BITS file. Because some establishments with mid-year reorganizations report March employment in both of their records, the aggregate employment from the BITS is slightly lower that that from the SUSB file for each year, since only the employment from the second record for an establishment identified as a midyear reorganization was used if that employment was non-zero.

The Company Statistics Division of the Census Bureau prepares for the SBA Office of Advocacy an extensive set of tables of aggregated BITS data on annual gross flows

(establishment and firm start-ups and closures, employment gains from start-ups and expansions of establishments, and employment losses from closures and contractions) for multi-unit and single-unit firms, by firm size, industry, and location. They also prepare custom tabulations on a contract basis for specific research projects. However, many research needs cannot be met by these tabular data designed for public use because they are limited to predefined cells representing some minimum number of firms, and do not allow exploration of the data, nor refinement of the specification of variables. Access to the microdata is also necessary to avoid biasing analyses as a result of the necessary suppression of any small cell values and their complementary cells during the disclosure processing for tabular data. These small cells may contribute crucial information to statistical analyses performed within a Census Center for Economic Studies, while avoiding any disclosure problems in the empirical results.

3) Literature review of the existing research using BITS

The original evaluation of the BITS database was carried out and documented by Acs and Armington (1998) and Robb (1999). The database was first used to measure job flow dynamics in the U.S. economy, investigating both gross and net job flows (Acs, Armington, Robb, 1999). Much of the previous work on U.S. job flows had been based on analysis of establishments in the manufacturing sector only (for example, Davis, Haltiwanger, Schuh, 1996, based on Census' Longitudinal Research Database of manufacturing establishments).

Two further early studies were carried out while the authors were ASA/NSF/CENSUS research fellows at the Center for Economic Studies (CES) at the U.S. Bureau of the Census, Washington D.C., under grant # SBR 980894. These studies compared gross job flow dynamics in services and manufacturing (Acs and Armington, 2,000) and examined the persistence of new jobs (Armington and Acs, 2004).

The survival of start-up business was examined in several studies using BITS (Headd, 2001). A second study demonstrating the analytical capabilities of the BITS data to explore firm survival analyzed tabulations for several industry groups—information technology, goods producing and service producing industries (Boden, 2000). A comparison of employment changes in acquired establishments with those in non-acquired establishments (Armington and Robb, 1998) showed that those acquired from small firms experienced more rapid job growth, while those acquired from large firms experienced more losses than gains.

Another paper dependent on the detailed establishment and firm microdata (Bjillalonga, 2001) examines whether the finding of a diversification discount in U.S. stock markets is only a data artifact. This study used BITS to construct business units that are more consistently and objectively defined across firms, and thus more comparable.

Two studies have examined the size distribution of firms. Mergers in economic concentration were examined by White (2002). White asked the question, "Has the wave of mergers in recent decades significantly changed the size distribution of large firms?" Are the largest firms in the economy growing relative to the overall size of the economy? The results show clearly that despite the recent waves of mergers, aggregate concentration declined during the 1980s and the early 1990s and then generally increased after the mid–1990s, but the levels at the end of the decade remain at or below the levels of the late 1980s or early 1990s, and well below the levels of the early 1980s.

Analyses of firm sizes have historically used data that included limited samples of small firms, and these have generally concluded that firm size distributions are typically lognormal. Using BITS plus data on firms without employees to represent the entire population of tax-paying firms in the United States (Axtell, 2001) shows that the Zipf distribution characterizes firm sizes. The probability that a firm is larger than size s is inversely proportional to s. These results hold for data from multiple years and for various definitions of firm size.

Analyses of cities, states and other regional units are another application benefiting from use of the BITS data. The impact of the unequal distribution of university research and development funding on regional firm formation and economic growth rates was analyzed by Kirchhoff *et al* (2002). Lee, Florida, and Acs (2004) explored whether connections exist among regional social characteristics, human capital, and new firm formation. They argue that social diversity and creativity have a positive relationship with new firm formation. Plummer and Acs (2004) employed spatial econometrics to examine the impact of new firm formation and knowledge on economic growth in Colorado. They find that new firms are more important to regional development than incumbent firms.

III. USING CENSUS'S BITS DATABASE TO EXPLORE ENTREPRENEURSHIP, GEOGRAPHY, AND ECONOMIC GROWTH

1) The theoretical framework

The knowledge-based growth models have three cornerstones: spatially constrained externalities, increasing returns in the production of goods, and deceasing returns in the production of knowledge (Romer, 1986, 1990; Lucas, 1988). New knowledge—in the form of products, processes or organizations—leads to opportunities that agents exploit commercially. Such opportunities are then a function of the distribution of knowledge within and between societies. But opportunities rarely present themselves in neat packages—they have to be discovered and packaged. Precisely for that reason, the nexus of opportunity and enterprising individuals is crucial in understanding economic growth (Shane and Eckhardt, 2003).

However, the ability to transform new knowledge into what Arrow (1962) designated 'economic knowledge' (leading to commercial opportunity) requires a set of skills, aptitudes, insights and circumstances that is neither uniformly nor widely distributed in the population. Moreover, empirical findings support the proposition that entry and entrepreneurship provide important links between knowledge creation and the commercialization of such knowledge, particularly at the early stage when knowledge is still fluid (Audretsch and Keilbach, 2004).

The basic shortcoming of the endogenous growth model is that it fails to recognize that only some of the aggregate stock of knowledge (often associated with R&D costs or products) is economically useful, and that even economically relevant knowledge may not be successfully exploited if the transmission links are missing. Furthermore, much of the general stock of knowledge is not in the public domain, and it may not spill over easily from one carrier to another. Most knowledge, regardless of whether it is in the public or private domain, requires a certain absorptive capacity on the part of the recipients in order for successful transmission to occur. This suggests that there is a filter between the stock of knowledge and the more limited economically useful knowledge. Not only does the level of knowledge vary among countries and regions; the transmission capacity of the filter also varies.

Consequently, despite the gains in terms of transparency and technical ease obtained by imposing strong assumptions in the endogenous growth models, these advantages have to be measured in relation to the drawbacks of deviations from real world behavior. Hayek (1945) pointed out that the central feature of a market economy is the partitioning of knowledge or information about the economy. The endogenous model fails to incorporate one of the most crucial elements in the growth process; transmission of knowledge through entrepreneurship—new firm formation—and the resulting spatial dimension of growth.

Thus, a closer connection between the endogenous growth models and the models of entrepreneurship seems necessary. The fact that knowledge-producing inputs are not evenly distributed across space implies that regions may not grow at the same rate, not only because they have different levels of investment in knowledge but also because they exploit knowledge at different rates. Even if the stock of knowledge were freely available, including the tacit and non-tacit parts, the ability to transform that knowledge into economic knowledge, or commercialized products, would not be. Moreover, most knowledge is not a free good at everyone's disposal. Often only a few individuals know about a particular scarcity, or a new invention, or a particular resource lying fallow or not being put to best use. This knowledge is both idiosyncratic and local, because it is acquired through each individual's own networks, depending on their occupation, on-the-job routines, social relationships, and daily life. It is frequently this particular knowledge, obtained through a local knowledge network, which leads to profit-making insight (Michelacci, 2003).

The dispersion of information among different economic agents who do not have access to the same observations, interpretations or experiences, has implications for economic growth.

Since this is not recognized in the endogenous growth model, we need to extend it with some additional assumptions and outline an alternative structure to improve the model. In order to remedy the limitations of the endogenous growth model and to specify the nature of the transmission mechanism that diffuses knowledge and converts it, via entrepreneurship, to growth, we propose the following assumptions.

1 New firms are assumed to be the primary mechanism to commercialize new knowledge, regardless of whether it is drawn from the stock of existing knowledge, or is newly discovered, and whether it is scientific knowledge or other. This transformation into economically relevant knowledge often occurs via spillovers that are exploited in new ventures, evidenced in firm formations. When existing firms acquire new economic knowledge, they may create new establishments within the firm, but the majority of such secondary new establishments are replications of other establishments owned by multi-location firms. Thus new firm formations are seen as the primary indicator of knowledge spillovers leading to economic growth.

2 Each new firm embodies a new idea, or innovation, expecting to provide a new, or improved, or more competitive products or services to customers. Schumpeter (1911)[5] suggests that a new idea (innovation) represents any kind of new combination of new or existing knowledge. These new firms are extremely heterogeneous, not only in the size, but also in terms of characteristics such as absorptive capacity, strategy, technology, product range, and performance (profitability, productivity, etc.). Because new entrants often make mistakes and fail, a very high formation (or gross entry) rate is necessary to sustain long-term growth.

3. Knowledge spillovers are primarily local events; there are few important interregional spillovers. Success in converting available public or private knowledge into economically useful firm-specific knowledge depends on the initiative and skills of the local potential entrepreneurs, and these entrepreneurial conditions vary across regions. Local policy and previous history (path dependence) determines the local entrepreneurial climate, which may be embedded in the local infrastructure, regulation, attitudes, educational policies, networks, technology transfer mechanisms, etc.

The combined result of these assumptions, when added to the endogenous growth model, can be characterized as a filter (here defined in terms of entrepreneurship) that determines the proportion of local knowledge that is converted into economically useful firm-specific knowledge (Acs, Audretsch, Braunerhjelm and Carlsson, 2003). This suggests that an increasing stock of knowledge (through R&D and education) will lead to higher economic growth only if the knowledge is economically useful and if the economy is endowed with factors of production that can select, evaluate and transform knowledge into commercial use, i.e., entrepreneurs. If these conditions are not fulfilled, an increase in the knowledge stock may have no impact on growth. Similarly, highly entrepreneurial regions with smaller knowledge stocks may experience higher growth than regions more abundantly endowed with knowledge.

The basic structure of the model accommodates both incumbent firms and new firms. Incumbent firms accumulate knowledge over their lifetime, and this accumulated firm-specific knowledge influences their ability to exploit new knowledge spillovers—the degree of firm specificity of their existing knowledge constrains their future absorption of knowledge spillovers. Hence, the incumbent firms' ability to exploit spillovers is determined by path-dependence. Furthermore, new establishments that are created by incumbent firms may be located in other regions, since their location decisions are likely to be based on cost-minimizing decisions.

New firms differ from incumbents, in that their economic knowledge is not governed by path-dependence to the same extent, but is built on the local entrepreneurs' ability to exploit opportunities arising from aggregate spillovers. Start-ups entering the market thus provide direct

evidence of the conversion of knowledge to growth. They produce genuinely new products and services, or compete using new processes or filling under-served niches.

Both types of firms exploit knowledge spillovers, albeit in different ways. Together their performance determines the share of knowledge spillovers that are commercialized. We can think of θ as the absorptive capacity of incumbent firms and λ as a proxy for entrepreneurship within an economy. Then, in accordance with assumptions 1 and 2, the standard production function has to be modified to account also for entrepreneurship:

$$F(k_i, x, (\lambda + \theta)K)$$

where k_i is new knowledge produced by firm i, *x* is a vector representing all other inputs, and since each individual firm cannot appropriate all the knowledge they produce in the first period, *K* is a pool of knowledge from which spillovers occur. Thus, if entrepreneurship is non-existent in an economy (so λ is zero) and θ is constant, then knowledge spillovers will not provide the same solution as in the endogenous growth model with automatic and all encompassing spillovers. In fact, the model will then reduce to the neoclassical growth model.

In addition, it is obvious that it is not only the size of *K* and the absorptive capacity of incumbent firms that matter, but also the presence of entrepreneurs as captured by λ. Our empirical work was driven by the effort to estimate the size of λ. In Romer's work (1990), λ equals unity, which implies that all knowledge (*K*) is accessible and convertible into economic knowledge (Acs and Varga, 2002), a very strong and unlikely assumption (Acs and Varga, 2004).

A rich literature exists in regional economics that sheds some light on how to capture the extent to which pooled labor markets, non-pecuniary transactions, and information spillovers exist. One approach suggests that the infrastructure of services is more developed in regions that are more densely populated. According to Krugman (1991a, p. 484), "The concentration of several firms in a single location offers a pooled market for workers with industry-specific skills, ensuring both a lower probability of unemployment and a lower probability of labor shortage." Thus the start-up rate for each industry sector should increase with the existing density of establishments in each sector. Another view is that localized industries tend to support the production of non-tradable specialized inputs. Thirdly, informational spillovers give clustered firms a better production function than isolated producers have. The high level of human capital embodied in their general and specific skills is another mechanism by which new firm start-ups are supported. Thus regions that are rich in this resource should have more start-up activity. University graduates—especially engineers—provide a supply of labor to local firms. New firm start-ups should be positively related to higher average levels of education, and negatively related to the levels of unskilled and semi-skilled workers in the region. In most studies of new firm formation in the 1980s there was a heavy emphasis on the explanatory power of unemployment (Evans and Leighton, 1990, Storey, 1991). It was suggested that when workers are unemployed they might be more likely to start their own businesses. The formation of new firms, in turn, may reduce the unemployment rate as such persons start new firm employing not only the owner, but also others. However, higher levels of unemployment might also indicate a reduction in aggregate demand throughout a regional economy, thereby putting downward pressure on the rate of new firm formation. Storey (1991, 177) found that, generally speaking, time series analyses point to unemployment being, *ceteris paribus*, positively associated with new firm formation, whereas cross-sectional, or pooled cross-sectional studies appear to indicate the reverse. This may be due to sectoral differences, with the industry sectors that require relatively small amounts of capital being more suitable for startups in periods of higher unemployment.

Also associated with studies of new firm formation from the 1980s was the role of industrial restructuring. Industrial restructuring has been associated with (1) the shift from manufacturing employment to services, (2) a reduction in both firm and plant size, and (3) a shift to higher levels of technology. The shift from manufacturing to services, which are usually less

capital intensive than manufacturing, could increase the rate of new firm formation. Regions that are dominated by large branch plants or firms will have less new firm formation (Mason 1994), in part because such areas have relatively fewer people with the managerial or skilled labor backgrounds that are the source of most firm founders.

2) The construction of variables for empirical examinations

A. The unit of observation for these studies

Although the BITS data support analysis at the firm level, these studies focus on the analysis of regional variations within the United States, and seek understanding of why various levels of economic activity vary across regions. Therefore, after considerable preliminary analysis of the data at the firm level, the scope and definitions of the relevant regional data were carefully defined and the firm-level data were aggregated to create regional data.

The geographic unit of analysis chosen for this study, the Labor Market Areas (LMAs) defined by Tolbert and Sizer (1996), is ideal for our purposes because it identifies economic areas broad enough to contain most of the labor supply and the local market for their business population, while being small enough to substantially avoid the worst of the aggregation problems of larger geographic units. These LMAs are aggregations of the 3,141 U.S. counties into 394 geographical regions based on the predominant commuting patterns (journey-to-work) between them in 1990. Each LMA contains at least one central city, along with the surrounding counties that constitute both its labor supply and its local consumer and business market. Many of the 394 LMAs cut across state boundaries, to better define regionally integrated areas of local economic activity.

Tolbert and Sizer specified these LMAs for the Department of Agriculture, using the Journey-to-Work data from the 1990 U.S. Census of Population. The LMAs are named according to the largest place within them in 1990. Some LMAs incorporate more than one MSA, while others separate some of the larger MSA's into more than one LMA, depending on the commuter patterns. A few smaller independent (usually rural) Commuting Zones have been appended to adjacent LMAs so that each LMA had a minimum of 100,000 population in 1990, which is necessary to avoid possible disclosure of confidential Census data that have been aggregated for LMAs. Alaska and Hawaii each are treated as a single integrated LMA, although they clearly have little mobility across their entire areas. See Reynolds (1994) for further discussion of why LMAs are the most suitable unit for this type of analysis.

The LMA unit of observation has the advantage of including both the employment location and the residence location of the population and labor force within the same area. Being based on counties, a wide variety of data collected at the county or ZIP-code level can be aggregated to construct LMA-level data. Finally, the 394 LMAs together cover the whole country, so that their data can be aggregated to U.S. totals, and all areas are represented.

B. Measurement of formation and growth rates

In the research reported here we are investigating regional differences in gross new firm formation rates, not the net change in numbers of firms or establishments in an area. The factors we are focusing on to account for differences in rates of new firm formation include local differences in educational attainments, entrepreneurship, innovation, and industrial evolution. The factors contributing to explanations of local differences in firm deaths, plant entry and exit, all of which affect the net numbers of establishments, are far beyond the scope of this paper, and generally not strongly related to local human capital.

Firm formation rates are calculated for each of the 394 LMAs, based on new firm formations during the period under study.[6] Single unit firm formations in year t are identified on the BITS as non-affiliated establishments that reported a Census start-year of t or t-1, that had no employment in March of year t-1, and had positive employment below 500 in March of year t.

(The Census start-year is the year that the establishment first reported any payroll and therefore entered the Census business register.) This avoids inclusion of either new firms that have not yet actually hired an employee, or firms recovering from temporary inactivity. 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 have no employees in March, but recover some employees the following year. Those new firms that had 500 or more employees in their first year of activity appear to be primarily offshoots of existing companies.[7]

New firm formations include most of the primary locations of the relatively few multi-unit firms (1500 to 6000 per year) that appeared to start up with less than 500 employees (firmwide) in multiple locations in their first year. We limited multi-unit firm formations to those whose employment in their new primary location constituted at least a third of their total employment in the first year.[8] 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.

Because the Labor Market Areas vary greatly in size, the absolute numbers of new firm formations must be standardized by some measure of the LMA size before it is meaningful to compare them across areas. When dealing with the whole service sector, firm formation rates are calculated as the number of new firms per thousand members of the labor force in the LMA in the prior year. This labor force approach has a particular theoretical appeal, in that it is based on the theory of entrepreneurial choice proposed by Evans and Jovanovic (1989). Each worker in the LMA chooses whether to be an employee of an existing business, or to become an entrepreneur and form a new firm. This approach implicitly assumes that the entrepreneur starts the new business in the same labor market where he or she previously worked or sought employment. It also has the added property that there is a clear lower bound of 0.00 (for no new businesses), and a theoretical upper bound of 1.00, which would represent the extreme case where every worker within a region started a new business during a year.

However, when comparing firm formation rates for different industries, or across subsectors of the service industry, we need to standardize for the differences in sizes of both areas and subsectors. For this purpose we calculate formation rates in terms of the number of new formations per thousand establishments already in existence in that industry or subsector in each LMA. This could be termed the ecological approach, because it considers the amount of start-up activity relative to the size of the existing population of businesses in that sector.

Two considerations of the timing of the firm birth rate data should be noted. While new firms enter the business register underlying the BITS file on a nearly continuous basis, their employment data are reported only for a pay period in March of each year. Since we require positive employment before recognizing a new firm, if a firm begins hiring after March, we do not count its formation until the following year. Therefore, each specified year's firm formation counts actually represent firms that hired their first employees sometime between April of the prior year and March of the specified year, for an average of nine months lagged reporting. Further, Reynolds et al (1995) and others have shown that the time between an individual's decision to create a new firm and the start of the resulting economic activity averages about two years, and is often longer. With such lags in the initialization and reporting of new firm formations, we would not expect to be able to identify a lag structure between differences in their annual rates and the regional factors associated with these differences, even though we have nine years of annual data on new firm formations.

C. Industry sectors

We distinguish six broad industry sectors for the analysis of growth, to facilitate analysis of different industries' sensitivities to factors affecting their growth, and to better control for aggregation effects in regions with different shares of weak industries—manufacturing,

agriculture, and mining sectors. This expands both the scope and the industrial detail beyond that of previous studies, most of which were limited to manufacturing. Industry codes are based on the most recently reported 4-digit SIC code for each establishment, because the precision and accuracy of the codes tends to increase over time.[9]

Sector	Standard Industrial Classifications
Distributive	4000-5199 (transportation, communication, public utilities, and wholesale trade)
Manufacturing	2000–3999
Business Services	7300–7399 and 8700–8799 (incl. engineering, accounting, research, and management services)
Extractive	0700–1499 (agricultural services and mining)
Retail Trade	5200–5999
Local Market	1500–1799 and 6000–8999 excl. Business services (construction, consumer and financial services)

These six broad sectors distinguish industries that might differ in their sensitivity to local market conditions. For instance, local consumer services and construction are more dependent on local regional demand than manufacturing and distributive services are, while manufacturing and distributive services may depend more heavily on the supply of semi- and unskilled labor. Growth in extractive industries is limited by the local supply of natural resources and arable land.

D. Variability in formation and growth rates in the 1990's

During the period from 1991 to 1996 U.S. private employment increased over 10%, while the employment gains from new establishments during that period contributed over 26%. Looking at the comparable growth rates for the six industry sectors we distinguished, in Table 1, it is apparent that the greatest growth was in business services (28.7% net increase, with 43.6% gross increase from employment in new establishments), followed by other local market services and construction (12.2%, with 25.8% in new establishments). Employment in manufacturing and extractive industries was virtually constant, although both showed substantial gains in new establishments, indicating that those sectors were continuing to evolve new products and processes to replace those that were discontinued or shrinking.

When we shift from the national totals to looking at the (unweighted) averages of our regional data for the 394 LMAs the mean growth and formation rates are less striking, but the variation across regions is remarkable. Average annual employment growth varies from a small loss to annual growth of 8%, while firm formation rates vary from 2 new firms per thousand labor force to a high of 10 new firms per thousand labor force. The local market sector mean firm formation rate accounts for nearly half of the total, although that sector accounted for just over a third of the employment. Similarly, business services, retail trade, and even the extractive industries accounted for somewhat more of the firm formations than their shares of employment. Firm formations in the manufacturing and distributive industries both lagged their employment shares. Employment in an area tends to keep pace with the growth of population in that area, *ceteris parebus*, so it is useful to examine both the rate of increase in employment, and how it differs from the rate of increase in population. It is not clear whether the growing economy is attracting the increasing population, or the growing population is simply causing the economy to expand to keep up with local demand and supply.

Table 1: Establishment employment, gross change in births, and firm formation rates 1991–1996, by firm type, and by industry sector

Establ. Class	Employment		91-96 Empl. change		LMA Employment growth ratios annualized			LMA Firm formation rate per 1000 labor force		
	1991	1996	net	birth	Mean	Min.	Max.	Mean	Min.	Max.
All	92,265,576	102,149,281	10.2%	26.3%	1.03	0.99	1.08	3.67	2.05	10.00
Firm type										
Single unit	38,532,294	44,811,609	15.1%	31.3%						
Multi-unit	53,731,429	57,324,994	6.5%	22.6%						
Industry Sector										
Bus. services	7,780,445	10,385,762	28.7%	43.6%	1.07	0.94	1.39	0.35	0.11	1.14
Distribution	11,887,375	12,719,155	6.8%	23.4%	1.02	0.95	1.10	0.41	0.20	1.72
Extractive	1,269,551	1,237,600	-2.5%	24.5%	1.01	0.79	1.27	0.09	0.01	0.51
Local market	33,434,183	37,773,144	12.2%	25.8%	1.04	0.98	1.11	1.75	0.94	5.20
Manufactures	18,450,502	18,556,546	0.6%	13.3%	1.01	0.92	1.13	0.19	0.06	0.50
Retail trade	19,443,520	21,477,074	9.9%	33.3%	1.03	0.99	1.09	0.88	0.52	2.61

Change rates are based on the mean of 1991 and 1996 employment for the class of establishments.
Formation rates are 1991–1996 annual average new firms in class per 1000 labor force in LMA in 1993.
Type = multi if establishment part of multi-unit firm in either year.

Source: 1989-96 LEEM file tabulations at Census' Center for Economic Studies, and Bureau of Labor Statistics.

IV. FINDINGS FROM ACS & ASSOCIATES USING THE BITS DATABASE

1) Regional variation in entrepreneurship[10]

A. Central hypotheses and model estimated

This analysis reexamines the issue of new firm formation in light of recent theoretical developments in economic geography and new growth theory, to the end of better explaining why new firm formation rates vary so widely across regions. A long tradition of studies of the location of new plant entry has focused on tax rates, transportation costs and scale economies at the plant level (Bartick, 1989). More recently, a growing literature has sought the determinants of variation in new firm formation on a regional basis (Reynolds 1994; Keeble and Walker, 1994; Audretsch and Fritsch, 1994; Reynolds, Miller and Maki, 1994). We focus primarily on four determinants of regional variation in the firm birth rate: (1) higher birth rates are promoted by regional spillovers, especially of relevant knowledge; (2) higher unemployment may deter start-ups in some sectors and increase them in others; (3) industrial restructuring should promote new firm formation; and (4) the existence of an entrepreneurial culture should promote start-up activity.

To test these hypotheses, we estimate a regression model where the dependent variable is the average annual firm birth rate in year t divided by the labor force in year t (in thousands). This is analogous to the method used by Keeble and Walker (1994).

B. The explanatory (exogenous) variables

Establishment size is a proxy for the structure of industry in the region. It is measured as employment in year t divided by the number of establishments in year t in the region. It should be negatively related to regional birth rate since larger average establishment size indicates greater dominance by large firms or branch plants.

In order to assess the potential for positive effects from spillovers, many studies have measured density using the square root of the regional population, or population per square mile. Such measures, however, do not indicate the extent of pooled labor markets very well, since they tell us nothing about the density of similar establishments in the region. Therefore, we introduce a new measure that captures both population density and the number of establishments in a region. *Industry density* is the number of establishments in the industry and region in year t divided by the region's population in year t. The greater the number of establishments relative to the population, the more spillovers should be facilitated (Ciccone and Hall, 1996).

Population growth is the average annual rate of increase in the region in a previous period (calculating the two-year change from the ratio of the population in year t divided by population in t-2, and taking the square root of that two-year change ratio to calculate the annual change ratio). *Income growth* is the average annual rate of increase of personal income in the region similarly calculated. Both of these growth factors from the period preceding our start-up measurement period are expected to promote new firm start-ups in the subsequent period.

The *unemployment rate* is the traditional calculation for the first year of our start-up measurement period—the average number of unemployed in year t divided by the labor force in year t. It is expected to be negatively related to start-ups overall, but probably positively related to new firm start up rates in industries with low capital requirements, and negatively related to those with high capital requirements. The simple correlation between the unemployment rate and the firm birth rate is close to zero, and is not statistically significant.

The *share of proprietors* in the economy is measured as the number of proprietors in year t divided by the labor force in year t. Proprietors are members of the labor force who are also business owners. It includes both the self-employed who have no employees, and the owners of unincorporated businesses that have employees. The simple correlation between the regional

birth rate and the share of proprietors is 0.30, indication a moderately strong positive relationship between these variables.

To measure the level of human capital in the economy we include two measures of educational attainment in each region. The first is the share of adults with *no high school degree*, defined as the number of adults without a high school degree in 1990 divided by the number of adults (population 25 years or older). The lack of a high school degree should be a good proxy for the proportion of unskilled and semi-skilled labor, and should be negatively related to the birth rate. The mean percentage of the population without a high school degree is 27%. In fact, the simple correlation between the percentage of the population without a high school degree and the birth rate is –0.19.

Finally, share of *college graduates* is defined as the number of adults with college degrees in 1990 divided by the total number of adults. This is a proxy measure of both technical skills needed in the economy, for example engineers and scientists, and skills needed to start and build a business, like finance and marketing and complex reasoning. In 1990 an average of 15.9% of the adult population had a college degree. Its simple correlation with the regional birth rates is positive.

C. Summary of results

Effects of regional characteristics on annual firm birthrates in LMAs is shown in Table 2 for 1990 through 1998. We present standardized beta coefficients, so that each parameter indicates the sensitivity of formation rate variation to normalized variation in the corresponding independent variable.[11] The estimated parameters are quite consistent across these years, and the signs are generally as expected. The effect of growth in per capita income decreased over time, becoming insignificant or tiny in the final three years. The unemployment rates had no significant impact in most years, but tiny positive impact in 1990 and 1994.

Table 3 shows the results of least squares regression on the 1994–1996 average annual firm formation rates for all industries together, and for each of the six industry sectors, based on the 394 Labor Market Areas. The t-ratios shown for each were calculated from the simple estimated standard errors, but they were very similar to those recalculated with a correction for heteroscedasticity. The estimated coefficients are generally consistent with our expectations, but with several important exceptions.

The explanatory and control variables together explained two-thirds of the regional differences in firm start-up rates for all industries together and for most of our six sectors. However, they explain much more of the local variation in formation of new business service firms, and much less of the variation for the distributive industries.

Industry density and population growth are both strongly positive and statistically significant, as predicted by the theory of regional spillovers (Krugman, 1991a and 1991b). In fact, these coefficients are nearly three times as large as the coefficient on income growth, which is also positive and significant. When analyzed separately for each of the industry sectors, we find that both industry density and population growth are positive and significant for each of the six industry sectors. However, the parameters on income growth are very small, and are only significant for business services and local market.

The coefficient for establishment size is negative and statistically significant for all industries, and for all sectors by business services and extractive industries, indicating that regions with predominantly smaller establishments have a higher start-up rate than regions with more large establishments. This supports the thesis that regions that have already restructured away from large manufacturing dominance have a higher start-up rate than regions that have not.

The coefficient for the unemployment rate is positive, although it is tiny and not statistically significant at the all-industry level. This result is surprising, given that previous cross-

sectional studies have generally found a consistently negative result (Storey, 1991). Furthermore, the coefficients on unemployment were positive for all of the six sectors, and significantly so for all but the extractive industries. Perhaps the exceptionally low levels of unemployment and even shortages of labor in the United States in the 1990's account for the prevailingly positive relationship between unemployment and new firm births in this period. The implication here is that as workers shift from being employed to unemployed, the overall entry rate in the region tends to go up slightly, although there is no evidence that it is necessarily the unemployed who are starting the new firms.

Finally, the coefficient for human capital, as measured by share of college graduates, is positive and statistically significant, suggesting that regions that have higher levels of education will have higher firm formation rates. This is consistent with Anselin, Varga and Acs (1997), who found that in technologically advanced industries individuals with greater skills, knowledge and expertise are more likely to start businesses. However, for both business services and manufacturing this coefficient is only barely positive, and not statistically significant. Reynolds (1994) found a negative and statistically significant relationship between college education and the new firm birth rate in manufacturing. The results do suggest that manufacturing firms may behave differently than other sectors of the economy.

Table 2: Effects of regional characteristics on annual firm birth rate of LMAs
(estimated standardized beta coefficients, with t-ratios below, bold if significant at .05 level)

	Year t, for firm births in year t+1								
Firm births(t+1)/LabForce(t)	1998	1997	1996	1995	1994	1993	1992	1991	1990
Adj. R sqrd	0.61	0.66	0.63	0.67	0.67	0.63	0.61	0.64	0.61
Establ size	**-0.33**	**-0.36**	**-0.33**	**-0.35**	**-0.33**	**-0.29**	**-0.33**	**-0.35**	**-0.37**
empl (t) / establ (t)	-7.91	-9.68	-8.72	-9.05	-8.64	-7.67	-8.52	-9.28	-9.41
Business intensity	**0.44**	**0.45**	**0.38**	**0.43**	**0.49**	**0.48**	**0.49**	**0.45**	**0.45**
establ (t) / population (t)	9.85	10.96	8.71	10.52	11.72	11.20	10.98	10.66	10.25
Income growth (per capita)	**0.09**	0.02	0.07	**0.13**	**0.18**	**0.20**	**0.24**	**0.23**	**0.16**
sqrt(pers.inc(t) /pers.inc(t-2))	2.39	0.45	1.44	3.20	3.87	4.35	5.83	5.91	4.06
Population growth	**0.46**	**0.50**	**0.46**	**0.43**	**0.36**	**0.37**	**0.32**	**0.32**	**0.31**
sqrt(pop (t) /pop (t-2))	11.87	12.89	10.43	11.47	7.92	7.88	7.50	7.86	7.20
Unemployment rate	0.08	0.04	0.03	0.06	**0.07**	0.01	-0.04	-0.01	**0.10**
unempl (t) / labor force (t)	1.86	0.97	0.70	1.61	1.99	0.27	-0.88	-0.23	2.25
No high school deg.(1990)	**0.23**	**0.26**	**0.20**	**0.22**	**0.26**	**0.22**	**0.26**	**0.22**	**0.23**
adults<hs /adults(25+yrs)	4.67	5.44	4.09	4.77	5.65	4.44	5.06	4.41	4.66
College grads (1990)	**0.26**	**0.27**	**0.27**	**0.27**	**0.26**	**0.18**	**0.20**	**0.19**	**0.22**
adult grads /adults(25+yrs)	5.32	6.03	5.86	6.21	5.97	3.88	4.26	3.96	4.49

n = 394 LMAs.
Firm births and establishments from 1989–1999 LEEM file, U.S. Bureau of the Census
Other LMA data aggregated from county data from Census and Bureau of Labor Statistics

Table 3: Estimates of effects of regional characteristics on firm formation rates of LMAs by sector
(estimated standardized beta coefficients, with t-ratios below, bold if significant at .05 level)

1994-96 firm births/LF	All ind.	Bus. serv.	Distribut.	Extract.	Local mkt.	Manuf.	Retail
R sqrd	0.67	0.86	0.49	0.67	0.65	0.63	0.64
Establ size	**-0.36**	-0.02	**-0.15**	-0.01	**-0.15**	**-0.22**	**-0.22**
ind. empl / ind. establ	-7.08	-0.82	-3.01	-0.37	-3.43	-5.48	-4.04
Industry intensity	**0.46**	**0.86**	**0.64**	**0.79**	**0.53**	**0.76**	**0.50**
94 ind establ / population	11.03	25.2	14.8	23.6	12.2	22.8	11.99
Income growth	**0.16**	**0.07**	0.08	0.01	**0.15**	0.03	0.09
sqrt(94pers.inc /92pers.inc)	3.31	2.33	1.44	0.25	3.14	0.68	1.75
Population growth	**0.37**	**0.17**	**0.27**	**0.18**	**0.40**	**0.15**	**0.32**
sqrt(94 pop /92 pop)	7.81	5.69	4.61	3.76	8.39	2.92	6.61
Share of proprietors	-0.01	0.03	-0.05	-0.01	**0.11**	**0.13**	**0.21**
94 proprietors / labor force	-0.29	1.14	-0.90	-0.20	2.66	3.34	4.73
Unemployment rate	0.08	**0.06**	**0.16**	0.07	**0.17**	**0.14**	**0.18**
94 unempl / 94 labor force	1.91	2.48	3.51	1.85	4.47	3.34	4.65
No high school deg.	**0.23**	0.03	**0.38**	**0.17**	**0.19**	**0.15**	**0.20**
90adults<hs /adults(25+)	5.09	1.05	6.85	3.84	3.83	3.11	3.90
College grads	**0.29**	0.01	**0.25**	**0.11**	**0.24**	0.08	**0.22**
90 adult grads /adults	6.36	0.33	4.30	2.46	4.82	1.58	4.44

n = 394 LMAs.
From 1989–1996 LEEM file at Census' Center for Economic Studies

The positive and statistically significant coefficient for the percentage of the population without a high school degree is at first surprising, but can be easily explained. The correlation between the share with no high school degree and the new firm start up rate is negative, –.19, as expected. However, it is much more strongly negatively correlated with college education, with a coefficient of –0.70. After controlling for the proportion of adults with college degrees, the additional effect of a greater share of less educated workers is to facilitate the start-up process by providing cheap labor for the new firms. This positive impact of 'no high school degree' after controlling for 'college degree' is consistent across most of the industry sectors, except for business services, with the strongest positive relationship appearing in the distributive industries. In summary, using annual data on firm births for 384 labor market areas, in six industry sectors, between 1991–1998, we find considerable variation in the new firm formation rate across regions, but little variation over time. Variations in the firm birth rates are substantially explained by regional differences in industry intensity, population growth and income growth, as suggested by the new economic geography.

2) Human capital and entrepreneurship[12]

A. Central hypotheses and model estimated

It is clear from the data shown earlier in Table 4 that the service firm formation rates vary greatly across local economic areas, and we will seek the determinants of this local variation in the same factors that contributed earlier to the explanation of the differences in formation rates for all industries. The agglomeration effects that contribute to new service firm formation can come both from demand effects associated with increased local population, income, and business activity, and from supply factors related to the quality of the local labor market and business climate.

Among areas with broadly similar regional demand and business climate characteristics, there are further differences in rates of new firm formation and economic growth that are associated with the specific qualities of their human capital, and the propensity of locally available knowledge to spillover and stimulate innovative activity which culminates in new firm formations. Highly educated populations provide the human capital embodied in their general and specific skills for implementing new ideas for creating new businesses. They also create an environment rich in local knowledge spillovers, which support another mechanism by which new firm start-ups are initiated and sustained. Thus, regions that are richer in educated people should have more start-up activity, and local new firm formation rates should be positively related to local educational attainment rates. Furthermore, areas that already have relatively intense development of service businesses are likely to have higher levels of service firm formations, resulting in large part from spillovers of relevant specialized knowledge. We would expect that areas with relatively high shares of high-school dropouts would have lower rates of new firm formation.

To test the basic hypotheses that the new firm formation rates are positively related to the level of human capital in a region, we estimate a regression model where the dependent variable is the average annual new service firm formation rate (dividing births by the labor force in thousands) for 1996–1998.[13]

B. The explanatory (exogenous) variables

The primary explanatory variables are the same measures of the level of human capital and knowledge spillover conditions that were used earlier. The share of adults that were *college graduates* in 1990[14] is a proxy measure that covers both technical skills needed in the economy, for example engineers and scientists, and skills needed to start and build a business, like finance and marketing and complex reasoning. Its simple correlation with the service firm formation rates in LMAs is 0.29 and we expect it to be positively related to the birth rate, even after controlling for other important factors (Glaeser et al, 1995; Rauch, 1993; Simon and Nardinelli, 2002).

The second measure of educational attainment that we use is the *high-school dropout* rate, defined as the percentage of adults (population 25 years or older) without college degrees who also do not have high school degrees in 1990. This high school dropout rate should be a good proxy for the proportion of unskilled and semi-skilled labor in the LMA, and should be negatively related to the birth rate. While many high-school dropouts are employed in some of the personal and business service activities, few of them have the skills to start and manage a new firm themselves. In fact, the simple correlation between the high-school dropout rate and the service firm formation rate is –0.19.

Formal education itself does not usually provide either the skills or the inspiration to start a new business. But higher education trains individuals to rationally assess information, and to seek new ideas. Therefore more educated people are more likely to acquire useful local knowledge spillovers from others who are involved in research or in managing some service business. The quantity of potentially useful knowledge spillovers is expected to be a function of the number of similar business establishments, relative to the population of the economic area. *Service-industry intensity* is defined as the number of service establishments in the region divided by the region's population in thousands. The greater the number of establishments relative to the population, the more spillovers should be facilitated due to density of establishments (Ciccone and Hall, 1996).

C. Summary of results

Table 4 shows the results of least squares regression on the 1990–1992, 1993–1995 and 1996–1998 average annual firm formation rates for the service sector for 394 labor market areas. We again present standardized beta coefficients, so that each parameter indicates the sensitivity of birth rate variation to normalized variation in the corresponding independent variable. The t-ratios shown for each were calculated from the simple estimated standard errors. These were also calculated with a correction for heteroscedasticity, and these results were found to be very similar to the uncorrected standard errors. The estimated coefficients are generally consistent with our expectations, but with several important exceptions. The explanatory and control variables together explain about two-thirds of the regional differences in new service firm formation rates.

Only two of the three human capital variables showed the hypothesized relationships. First, for human capital measured by share of college graduates, the coefficients are positive and statistically significant for all except the 1993-95 periods, confirming that regions with higher shares of college-educated adults have higher firm formation rates. This positive result on human capital is consistent with previous research (Storey, 1994).

The positive and statistically significant coefficient for high-school dropouts as a share of the non-college adult population is at first surprising—however it is consistent with our earlier results for the whole economy (Armington and Acs, 2002). There we suggested that after controlling for the proportion of adults with college degrees,[15] the additional effect of a greater share of less educated workers is to facilitate the startup-up process by providing cheap labor for the new firms. Even the most sophisticated businesses need some workers who are less educated to do the manual labor. Thus, the relationship between educational attainment and new firm start-ups at the regional level may be U-shaped, with both low levels and high levels of education conducive to firm formation and growth.

Thirdly, the coefficient on intensity of service establishments is positive and statistically significant, suggesting that regions that already have a relatively strong supply of service establishments will have higher rates of new service firm formation, as predicted by the theory of regional spillovers (Jovanovic and Rob, 1989). Indeed, this factor has the strongest relationship of any of our independent variables. The 0.60 value estimated for the standardized coefficient indicates that a locality with a service establishment intensity that is one standard deviation more intense than the mean will be likely to have firm startup rates that are 0.6 standard deviation higher than the mean.

Note also among the control variables that the coefficient on the unemployment rate is positive and statistically significant for 1990–1992, when the economy was undergoing a small recession. However, it is negative and barely significant during 1993–1995 and insignificant during 1996–1998, suggesting that this positive effect disappears as the economy improves, or as mean unemployment falls.

Table 4: Regression Coefficients for Service Firm Formation Rates in Labor Market Areas during three consecutive time periods**

(standardized betas with t-ratios below, significant at .05 or better unless starred*)

	1996-98	1993-95	1990-92
Adj. R sqd	.718	.658	.625
Human Capital			
College degrees / adults '90	**0.16**	**0.10***	**0.19**
	3.39	1.79	3.40
High-school dropouts / non-college adults '90	**0.16**	**0.21**	**0.14**
	4.21	4.86	3.20
Intensity of service estab/population	**0.63**	**0.60**	**0.47**
	6.53	5.63	4.26
Regional Characteristics*			
Population growth	**0.51**	**0.46**	**0.41**
	18.05	14.44	11.46
Per capita income growth	**0.09**	**0.19**	**0.13**
	3.03	5.62	3.77
Population (logarithm)	**0.22**	**0.16**	**0.18**
	4.92	3.25	3.6
Unemployment rate	**0.06***	**-0.09**	**0.17**
	1.64	-2.32	4.08
Avg. size of all establ (employment)	**-0.34**	**-0.33**	**-0.32**
	-8.06	-8.13	-7.30
Intensity of all estab/population	**-0.21**	**-0.07***	**-0.03***
	-2.42	-0.71	-0.31
Number of observations	394	394	394

** Formation rates are 3-year average formations per 1000 labor force in prior year

* Undated exogenous variables represent prior year, or prior 2 year averages

Hoping to better distinguish the impacts of our independent variables on the startup rates of various types of service activities, but limited by data disclosure constraints, we defined 9 service subsectors, using two dimensions that should be relevant to our analysis of variation in startup rates. The dimensions chosen were the market segment served and the customary education requirement for founder of new firms in each class of service activity. Each of these dimensions was broken into just three categories, so that applying both dimensions resulted in the classification of all services into nine subsectors, within which the service activities were fairly homogeneous with respect to these two dimensions.

We distinguish activities that are most frequently started by people who do not have college degrees (called 'high school' level for simplicity), from those generally requiring an 'advanced' (graduate, post-graduate, or professional) degree, and assigned the remainder to 'college.' These allocations were based on subjective judgment, using our general knowledge of service industries, supplemented by detailed descriptions of the 4-digit SIC classes in the 1987 Standard Industrial Classification Manual.[16]

Table 5 shows average annual firm formation rates and shares of total service employment for the nine services subsectors defined according to their market segments and founders' education requirements. [17] For the nine subsectors defined by the education requirement and the market segment together, the firm formation rate was highest, at 14.78, for businesses in non-local markets with founders normally having advanced degrees. The subsector requiring the same advanced degree for founders, but serving the local consumer market, had only 5.31 new firms for each hundred existing establishments in that subsector. For businesses that normally require a college degree for their founder, the birth rate is quite similar across all three of the market segments. Businesses requiring less educated founders (high school degree) also showed great variation across market segments, with high formation rates for non-local market, and low ones for the local consumer market.

Service firm birth rates were calculated as before for all firms for each of the 394 LMAs, based on new firm formations during each of three recent time periods—1990 through 1992, 1993 through 1995, and 1996 through 1998, and standardizing across different sizes of LMAs by dividing by the size of the labor force (in thousands) in the LMA in the prior year. However, for comparing new firm formation rates for different subsectors of the service industry we need to standardize for the differences in size of both areas and subsectors. For this purpose we express new firm formation rates in terms of the number of new firms relative to the number of establishments already in existence in that subsector and LMA. This could be termed the ecological approach, because it considers the amount of start-up activity relative to the size of the existing population of businesses.

In order to allow for variation in the estimated coefficients of variables that should be sensitive to our subsector dimensions, while controlling consistently for other regional characteristics, we expand the independent variables to be subsector-specific for the dimensions we want to test. Naturally, we expected the educational attainment variables to be sensitive to the Education requirement dimension.

Using the previous notation, this more detailed pooled estimation model has the following form:

(2) $\text{Birth rate}_{LEM} = f(\text{Coll}_{LE}, \text{HighSch Drop}_{LE}, \text{Subsector estab intensity}_{LEM}, \text{Pop gro}_{LM}, \text{Income gro}_{LE}, \text{Pop log}_{LEM}, \text{Unempl}_{LE}, \text{Estab Size}_{LM}, \text{All-ind estab intensity}_{LM})$.

In order to estimate this model, we first standardized all of the exogenous and endogenous variables to have a mean of zero and a standard deviation of one, within each of the nine subsectors. Therefore, each represents a relative measure for the LMA, within the subsector. Then we created dummy variables for each of the three values for each of the subsector dimensions—Market and Education. Finally, we multiplied each exogenous variable

times the appropriate dummies to create specialized exogenous variables that distinguished among the dimensions we wanted to test for differences in estimated coefficients. The results of the estimation of this model are shown in the last six columns of Table 6.

Table 5: 1996-98 Firm Formation Rates and Relative Size of Service Subsectors defined by Market Segments and Founder's Education Requirement

Education Requirement and Market Segment	Average annual* firm births per 100 estab in subsector		Share of services 1995 employment	
All services	8.84		100.0%	
All education classes				
Local business market		10.66		26.1%
Local consumer market		7.18		54.9%
Non-local markets		12.66		19.0%
High school				
All markets	9.29		30.4%	
Local business		12.22		9.3%
Local consumers		8.42		15.9%
Non-local markets		7.86		5.2%
College degree				
All markets	9.25		26.1%	
Local business		8.60		10.2%
Local consumers		9.08		13.6%
Non-local markets		10.72		2.3%
Advanced degree				
All markets	8.33		43.5%	
Local business		10.31		6.6%
Local consumers		5.31		25.4%
Non-local markets		14.78		11.5%

* The sum of firm births in 1996, 1997, and 1998 is divided by 3, and then divided by the number of establishments in 1995, and then multiplied by 100.

Table 6: Standardized* Regression Coefficients for Service Sub-sector Firm Formation Rates in Labor Market Areas during 1996-1998, with Subsectors defined by Education Requirement and Market Segment**

(betas with t-ratios below, significant at .05 unless starred*)

		Pooled, with Dummy-distinguished exogenous variables					
	Pooled 9	Educational Requirement			Market Segment		
	Subsectors	High School	College Degree	Advanc'd Degree	Local Business	Local Consumer	Non-local
Adj. R sqd	.566			.660			
Human Capital							
College degree/ adults '90	**0.14**	**0.01***	**0.21**	**0.11**			
	7.47	0.41	8.09	4.02			
High-school dropout / non-college adults '90	**0.09**	**0.03***	**0.04***	**0.16**			
	5.77	1.26	1.91	6.85			
Intensity of subsector establishments / popul.	**0.54**				**0.53**	**0.25**	**0.77**
	36.40				18.55	9.25	32.05
Regional Characteristics							
Population growth	**0.36**				**0.35**	**0.44**	**0.20**
	32.09				19.43	22.83	11.19
Per capita income growth	**0.07**	**0.06**	**0.08**	**0.06**			
	5.96	3.35	4.44	3.28			
Population (logarithm)	**0.16**	**-0.06**		**0.06**	**0.23**	**0.20**	
	9.18	-2.35		2.18	7.81	7.50	
Unemployment rate	**0.05**	**0.01***	**0.05**	**0.02***			
	3.23	0.47	2.54	1.15			
Avg. size of all establ (employment)	**-0.26**				**-0.17**	**-0.47**	**-0.14**
	-16.04				-6.87	-19.79	-6.36
Intensity of all industry establishments / popul.	**-0.06**				**-0.01***	**-0.02***	**-0.07**
	-3.59				-0.22	-0.74	-2.96
Number of observations	3546			3546			

** All variables are expressed as deviations from mean values for all LMA's within each subsector, and subsector firm formation rates are 3-year average formations per 1000 labor force in 1995.

Looking first at the human capital variables in these estimated models for subsectors in Table 6, we see that the share of adults with college degrees is not significant for the formation rate of service businesses requiring only a high school education for the founder. For services businesses requiring a college education, the variation in the local formation rates is much more sensitive than was indicated by either the all-service regression or the pooled subsector regression. There is also a significant positive relationship between the share of adults with college degrees and the formation rates of service businesses normally requiring an advanced degree for the founder. This results from the high correlation of the distribution of college degrees with that of advanced degrees.

The positive and statistically significant coefficient for the relationship of shares of high-school dropouts to formation of new service firms that require advanced degrees suggests that such businesses may be more dependent on having a large pool of unskilled labor. The statistically insignificant coefficients for the impact of the share of high school dropouts on formation rates in the subsectors of services that require only high school or college degrees suggests that such businesses are not as sensitive to the supply of unskilled labor. They may find that the unskilled labor supply in most areas is adequate for their needs.

The relative intensity of establishments in the same subsector of services is a significant explanatory variable for all market segments, but the formation of new firms serving non-local markets is particularly sensitive to the prior existence of similar businesses. This corroborates the many prior case study analyses that addressed the spillover effects of certain rapidly growing local industry clusters, and suggests that these spillover effects are particularly important for businesses that are not focusing on local markets.

Most of the estimated coefficients for regional characteristics crossed with education or market dummies were similar to those estimated without such distinctions. However, the differences that appeared are quite illuminating. The log of population was crossed with all six dummies, and the tiny and insignificant variables crossed with college degree and with non-local markets were later omitted, to strengthen the remaining estimates. These show that, unlike services to local markets, those to non-local markets are not sensitive to the size of the local economic area. Perhaps the high coefficient on subsector intensity for non-local services has captured all of the relevant agglomeration effects for that subsector. Distinguishing by the education dimension, larger population contributes a bit to the formation rate of services firms requiring founders with advanced degrees, but it reduces the formation rate of firms normally started by high school dropouts.

The coefficient on unemployment is positive and statistically significant only for service firms normally started by college graduates. This provides some clarification of the conflicting results found in previous studies of the effects of unemployment levels on new firm formation rates. Apparently, after controlling for regional differences in income growth rates, an increase in unemployment tends to lead to an increase in new firm formation by those with college degrees, but not by high school dropouts or those with advanced degrees.

Finally, the negative coefficient on average size of local businesses is strongest for formation of new firms serving local consumer markets, while that on the intensity of all establishments is significant only for formation of new firms serving non-local markets.

In summary, these results suggest that the regional differences in new firm formation rates do indeed depend to a large degree on the educational requirements and the market served by the newly formed firms. In particular, the local levels of educational attainment impact primarily the firm formation rates of the types of firms that are normally founded by better-educated entrepreneurs, and do not affect startup rates for those normally founded by individuals with less than a college degree. While formation rates of all service businesses are higher in areas with higher intensities of similar service establishments, new formations of firms serving non-local markets are three times more sensitive to this than those serving local consumer

markets, and those serving local business markets are twice as sensitive as those serving local consumers.

3) Employment growth and entrepreneurship[18]

A. Central hypotheses and model estimated

Our major hypotheses concerning the regional variation in the employment growth rates are related to dynamic externalities, and we capture the effects of these spillovers by analyzing how the employment growth rates vary across regions. The literature suggests that higher employment growth rates should be associated with increased entrepreneurial activity, with increased industry diversity and with higher levels of human capital.

We estimate a regression model where the dependent variable is (compounded) average annual employment growth rates over the 5-year period of the early nineties. This is measured as the fifth root of the ratio of 1996 employment to 1991 employment in each LMA and sector: The annual average growth rate of each local economic area is defined as:

$$\text{Average annual employment growth rate}_{srt+1} = (\text{empl}_{srt+1} \ / \ \text{empl}_{srt}) \ ** \ 0.2$$

For all industries together the local growth rates varied from 0.988 (or –1.2% annual average change) to 1.080 (or 8.0% annual average change). The equations are estimated for 394 LMAs for all industries together, as well as for each of our six industry sectors separately.

B. The explanatory (exogenous) variables

The flow of entrepreneurial activity is measured as the firm formation rate, including both new single unit firms (establishments, or locations) with less than 500 employees, and the primary locations of new multi-unit firms with less than 500 employees firm-wide (Armington and Acs, 2002).[19] Labor force is preferred to population or employment as a size indicator, because it is a better measure of the number of potential entrepreneurs in the local economy. Thus the birth rates are calculated as the number of new firms per thousand of labor force in each LMA.

The average birth rates for the period from 1991 to 1996 were calculated from the average of the number of births in 1992, 1993, 1995, and 1996,[20] divided by the labor force in 1993 in thousands.

Our second measure of entrepreneurial activity measures the business-owner share of the labor force in the area. This is not so much a measure of entrepreneurship as a measure of the local dominance of small business. Proprietors are members of the labor force who are also business owners. In addition to those who own firms with employees, this measure includes the self-employed who have no employees. The share of proprietors is defined for each LMA as the number of proprietors in 1991 divided by the 1991 labor force. This share averaged 20.5% nationally, and varied from a low of 9.9% to a high of 44.8% across LMAs.

We include two measures of agglomeration effects that characterized local economies. Therefore, we measure specialization as the industry intensity—the number of establishments in each industry and region in 1991 divided by the region's 1991 population. After standardizing by the national average, this measure is almost identical to the specialization measure used by Glaeser et al (1992). Industry intensity should be positively related to employment growth if specialization is important for regional growth.

To control for the vast differences in the physical density of economic activity we use *establishment density*, defined as the number of establishments per square mile in that industry in 1991. Establishment density should be positively related to local growth rates if agglomerations drive demand or increase network externalities (Ciccone and Hall, 1996).

We include two measures of human capital that have been found to have a positive impact on regional growth in previous studies (Simon and Nardinelli, 2002). The first is the share of adults with at least a high school degree, with adults defined as persons 25 years or older. Those adults without high school degrees are the principal supply of unskilled and semi-skilled labor for work in manufacturing branch plants and retail or unskilled service establishments. Higher shares of high school graduates indicate a generally higher level of human capital in the area. In 1990 73.0% of adults had at least a high school degree, nationally.

The second measure of educational attainment is the share of college graduates, defined as the number of adults with *college degrees* in 1990, divided by the total number of adults. We expect that employment growth will be positively related to higher average levels of education, at both the high school and the college level (Glaeser et al, 1995).

To control for differences in the size distribution of businesses in each industry and region, we include average local *establishment size*, measured for each industry sector and economic area by dividing the number of local employees in 1991 by the number of local establishments in 1991 in each sector.

C. Summary of results

We looked first for evidence of a lagged impact of our explanatory variables on the regional growth rate. Because of the relative stability of the explanatory variables over the decade of the 90's and the limited amount of employment data available by LMA, we chose to calculate average 3-year growth rates for 3 consecutive periods, and test which of the 3-year growth rates was best explained by the regional characteristics from the beginning of the 90's.

Table 7 shows the results of these regressions, with the regression on growth from 1990 to 1993 explaining 39% of the regional variation. This was much higher than for any of the longer lagged periods, so we felt that it was most appropriate to examine growth rates without the lags that some theories predict. Therefore we used contemporaneous 1991-1996 firm formation rates as the primary predictor of regional differences in the 1991-1996 growth rates for our analysis by sector.

There are three important results in the estimated model of local growth differences by sector that are presented in Table 8. First, the coefficient on the firm birth rate, which serves as a proxy for entrepreneurial activity, is positive, large, and statistically significant, as hypothesized. These results are robust for five of our six industry sectors, with the exception of manufacturing, where it was insignificant. This exception explains the prior findings of industrial organization economists that entry is not important for employment growth in manufacturing (e.g. Geroski, 1995). Much of the research in industrial organization, labor economics and regional science has been limited to analysis of data from the manufacturing sector, and these results have been frequently generalized to the whole economy. It appears that those generalizations from the behavior of manufacturing firms are not always valid, but they may be valid for other industries dominated by large plants.

The coefficient on the share of proprietors is positive and statistically significant suggesting that the greater the share of proprietors in a region the higher the growth rate. However, this relationship did not hold up for most of the industrial sectors, probably because the appropriate sector-specific data were not available for share of proprietors. The coefficient for the share of proprietors is only about one quarter of that for entrepreneurial activity, suggesting that it is not so much the accumulated stock of entrepreneurial activity but the flow that is important for economic growth. This result also suggests that it is younger age and not smaller size *per se* that is more important for promoting growth and productivity.

Second, the negative and statistically significant coefficient on industry intensity suggests that greater geographic specialization (or less industrial diversity) leads to less growth, rather than greater growth. These results are again robust for all industries sectors with the exception of

manufacturing, where the coefficient is positive but not significant. This suggests that specialization does not generally lead to higher levels of technological externalities or other knowledge spillovers that promote growth in the same industry sector. This is consistent with the findings of Glaeser et al (1992), Feldman and Audretsch (1999) and Acs, FitzRoy and Smith (2002).

The negative and statistically significant coefficients on establishment density suggest that when other factors are the same, employment growth will be greater in regions that have less physical crowding in their industry. Thus, when measured by the number of establishments per square mile, the agglomeration effect on growth seems to be negative for Labor Market Areas. This is in contrast with the findings of Glaeser et al (1992) and Ciccone and Hall (1996), who used growth in other industries in each area as an indication of the size of the agglomeration effect, and found a positive relationship with growth. Indeed, it contrasts with much of the theoretical literature on agglomerations (Krugman, 1991).

Third, the greater the proportion of the area's adults with a high school degree, the higher the growth rates. However, after all of the other exogenous variables are taken into account, the additional impact of higher proportions of college graduates was negative but insignificant. These results suggest that a broad basically educated labor force may contribute more to growth than the presence of relatively large numbers of college graduates. These results are consistent with Glaeser et al (1995) and Simon and Nardinelli (2002). These human capital variables were weaker and inconsistent for the various industry sectors. When the all-industry regression was run without the college graduate measure, the results were virtually unchanged (not shown). Both of these human capital variables were dropped and this had no substantial impact on the estimated parameters for the remaining variables either. Therefore, the results are robust with respect to the inclusion or exclusion of the human capital variables.[21]

In summary, higher rates of entrepreneurial activity in an industry sector and region imply lower barriers to birth and greater local competition, so this analysis can also be interpreted as an investigation of the impact of local competition on local economic growth. We found that higher rates of entrepreneurial activity were strongly associated with faster growth of local economies.

Table 7: Analysis of Local Factors Associated with Differences in Lagged Employment Growth Rates in LMAs
(estimated standardized beta coefficients, with t-ratios below, bold if significant at 0.05 level)

3-year empl change rate	1990-93	1993-96	1996-99
Adj. R sqrd	0.397	0.215	0.282
Observations	394	394	394
Entrepreneurial activity	**0.498**	**0.535**	**0.516**
avg ann formation rate 1991-93	9.47	8.92	9.00
Share of proprietors 1990	**0.273**	0.081	0.080
proprietors / labor force	4.44	1.15	1.19
Business specialization 1990	**-0.382**	**-0.217**	**-0.347**
establishments / population	-6.17	-3.07	-5.13
Business density 1990	**-0.411**	**-0.244**	0.076
ln (estabishments /sq miles)	-6.76	3.51	1.15
Basic human capital 1990	**0.144**	**0.170**	**0.235**
high-sch. degr./adults (25+yrs)	2.32	2.39	3.46
Higher human capital 1990	-0.056	-0.023	**0.176**
college degr./adults (25+yrs)	-0.88	-0.32	2.53
Establishment size 1990	**0.255**	**0.347**	**0.146**
employment / establishments	4.00	4.77	2.10

Note: Employment derived from Census' County Business Patterns.

Table 8: Analysis of Factors Associated with Differences in Employment Growth Rates in LMAs by Industry Sectors

(estimated standardized beta coefficients, with t-ratios below, bold if significant at 0.05 level)

1991-96 empl change rate	All ind.	Bus. serv.	Distribut.	Extract.	Local mkt.	Manuf.	Retail
R sqrd	0.33	0.10	0.13	0.31	0.44	0.14	0.25
Observations	394	394	394	394	394	394	394
Entrepreneurial activity	**0.62**	**0.57**	**0.41**	**0.42**	**0.54**	-0.04	**0.54**
avg ann 91-96formations/ 93LF	11.3	4.96	6.86	5.19	11.3	-0.59	8.98
Share of proprietors	**0.16**	**-0.14**	0.01	0.01	0.03	**0.21**	0.02
91 proprietors / labor force	2.81	-2.54	0.10	0.09	0.61	3.79	0.34
Specialization	**-0.30**	**-0.57**	**-0.28**	**-0.53**	**-0.50**	0.14	**-0.34**
91 ind establ / population	-4.78	-3.97	-4.10	-6.30	-8.40	1.92	-5.19
Density	**-0.22**	-0.05	**-0.11**	**-0.14**	**-0.13**	**-0.15**	**-0.23**
91 ind. estab /sq mile	-4.61	-0.91	-1.97	-2.84	-2.91	-2.75	-4.78
Human Capital	**0.14**	0.12	**0.16**	0.03	-0.08	0.06	-0.04
90 h-s degr./adults (25+)	2.02	1.61	2.06	0.40	-1.23	0.87	-0.56
Human Capital	-0.11	-0.03	0.06	0.13	0.10	**-0.15**	0.02
90 college degr./adults	-1.66	-0.34	0.73	1.87	1.51	-2.02	0.22
Establishment size	**0.20**	**-0.17**	-0.07	**-0.41**	**-0.25**	-0.09	0.07
91 ind. empl / ind. establ	3.26	-3.02	-1.03	-9.30	-4.55	-1.61	0.96

V. RELATIVE STRENGTHS AND LIMITATIONS OF BITS

1) Other longitudinal U.S. business microdata

Most business data sources that might be used for measuring new business formation are based on administrative data collected in conjunction with a government mandate to regulate or tax broad populations of businesses, or on commercial business data collected primarily for credit rating or marketing purposes. Thus each data system was designed for other purposes than statistical analysis of changes over time, and the decisions made in designing the collection, processing and organization of the data affect their suitability for such statistical analysis.

The U.S. Bureau of Labor Statistics (BLS) has broadly similar data on most businesses with employees, which is collected by states for administration of the Unemployment Insurance (UI) system. Each state forwards quarterly data from its databases for tracking the payments of UI taxes, along with supplementary data on most secondary locations within the state. These data are edited by the BLS and matched across time to build their longitudinal data base with quarterly reporting of employment and payroll. In spite of considerable investment by BLS to improve consistency of reporting across all the states, there are some underlying differences in coverage of the UI system (especially for non-profit firms), and no capacity to link all of the establishments belonging to any multi-location firm that operates across state lines. They could use the federal tax identification number (TIN) to link all establishments that belong to the same tax-paying entity, but this does not provide a linkage to the enterprise level for many of the multi-location firms. A substantial portion of the new businesses appearing in a state may be new secondary locations of existing businesses in other states with different TINs, so these data cannot be used for analysis of either new firm formations or employment dynamics of firms of different sizes.

The Census Bureau's Center for Economic Studies has been working for years on developing a linkage system for their archived historical files of SSEL data from 1975 to the present (Jarmin and Miranda, 2002). The SSEL has a broader scope than the County Business Patterns that serves as the base for the BITS, but its scope was expanding during the 70's and 80's. Further, much of the raw data is missing for several years in the '70s and '80s, adequate data for linking 1975 to 1976 are not available, and they need an to use an additional future year of data for refining the most current linkages. Therefore they have limited their longitudinal data base to data identifying establishment records and tracking them from 1976 through 1999 (as of July 2004). Estimates have been used to fill in much of the missing data (including all of 1978), and much editing has been completed to help identify births and deaths of establishments. However, work is continuing on the challenges of better identifying the firms to which these establishments belong, and of reducing the impact of Census processing cycles on the reported business populations and their attributes. Census's longitudinal data base project may eventually result in a longitudinal establishment and enterprise microdata file of greater scope and historical coverage, but it may also prove to be too complex a challenge to master for all but narrowly defined special industry studies.

The best-known commercial source of microdata on U.S. businesses is the Dun & Bradstreet (D&B) Corporation, whose Duns Marketing Indicators files have been linked over time to produce longitudinal data for tracking new firm formation and employment changes in establishments and firms. These data are collected primarily for credit rating purposes and are therefore continuously updated for the most active firms, and irregularly updated for many of the smaller firms. Therefore analysts have historically used end-of-year files, either annually or every second year, to represent annual (or biannual) data for measuring changes in the business population. The scope of coverage includes many government entities, and the universe has expanded over time, causing great difficulties for researchers who must edit it for consistency of coverage before using it to measure real changes in the economy. It also includes many self-employed unincorporated individuals as single-person firms, and includes partners in partnerships as employees. Many secondary locations of multi-unit firms are omitted, and some subsidiaries and businesses undergoing restructuring are double-counted. Nevertheless, with

careful editing these commercial data have been used to construct longitudinal microdata that showed the same patterns of change that are shown in the administrative data from the federal government (SBA Office of Advocacy, 1988a and 1988b). While these D&B-based data are not subject to the strict confidentiality requirements of the government data, they are very expensive to lease, and even more expensive to edit properly for research purposes. They also fail to identify many of the smaller new businesses until they grow big enough to apply for loans or accounts with large suppliers.

There are now other commercial sources of data on U.S. businesses, which generally derive their databases from computerized aggregations of Yellow Page entries or other secondary sources. Some of these use sophisticated matching software to remove duplicate entries, and merge in supplementary data, but none can rightfully claim to have either an economically well-defined universe or accurate employment data. Nor can any of these detect new firm formations promptly, because the U.S. does not have a national system for public registration of businesses, and increasing numbers of businesses do not bother with Yellow Page subscriptions, but rely on the internet, word of mouth, or private networks to attract customers.

2) Foreign data used for similar research abroad

The BITS data are much closer to the ideal data for analysis of entrepreneurship over time and space than the data available for most other countries. The business data maintained by Statistics Canada have long been the best example of comprehensive, longitudinal, business microdata, but they have not been prominent in analysis of geographic differences in entrepreneurship.

Most of the developed countries of Europe have been the subject of at least exploratory studies of their entrepreneurship rates and how they relate to their economic growth rates. The Global Entrepreneurship Monitor (GEM) project has been collecting comparable survey data on entrepreneurship from both developed and developing countries (40 countries in 2003), and these have been widely analyzed to show some of the impacts of entrepreneurship on growth. The GEMS data are limited to annual sample surveys of potential entrepreneurs, but they include an extensive list of variables relevant to analysis of entrepreneurship.

A number of international organizations have attempted to develop comparable data on entrepreneurship and growth for the various countries of Europe, by adjusting national data to improve consistency across countries. EIM in the Netherlands has a long history of working in this field, and now has produced their COMPENDIA 2000.2 (COMParative ENtrepreneurship Data for International Analysis), with harmonized data on the number of business owners in each of 23 OECD-member countries for 1972 to 2000. These data are derived from the OECD Labor Force statistics, but have been adjusted to standardize their scope and their definition of self-employment (or business-ownership) across countries, and to adjust for trend (coverage) breaks over time. However, these data show entrepreneurship only as the net changes in numbers of business owners. Therefore the actual new firm formation numbers are reduced by both the firm closures and the mergers, acquisitions, and sales of successful firms, leaving a residual that is very difficult to analyze.

For individual country studies a wide variety of indicators have been adapted as proxies for entrepreneurship, or rates of new firm formation by region (and/or sector). In addition to statistics on the actual numbers of new firms or establishments registered under various tax systems (such as value added tax in the U.K., German Social Insurance Statistics), other data on startups have been taken from the ZEW foundation panels that are based on data provided biannually by Creditreform, the largest German credit-rating agency, and from the Dun & Bradstreet files for other countries.

For other countries, the available data are more limited, either in scope or in level. The U.S. data cover all enterprises with employees (excluding the self employed unincorporated). The most comparable UK data cover all enterprises registered for VAT, and new entries are often

the result of changes in VAT requirements, or whether firms are subject to VAT. The Swedish data used for similar research cover only the manufacturing sector, whereas the German data have plants as their basic unit of account. Analyzing changes in these data sets over time and space inevitably provides a picture that is influenced by both the characteristics of the unit of measurement and the universe, as well as "real" changes in the economy. Perhaps in a further ten years a more consistent measures of enterprise and entrepreneurship will be available. On the other hand, increased globalization may simply make the problems of defining "new" firm formations and attributing employees to specific firms within a country more complex.

3) Limitations of BITS data

The scope of the BITS data is that of Census's County Business Patterns since 1989, and it therefore excludes agricultural production, forestry and fisheries, railroads, and domestic service workers from its coverage of private sector firms with employees. It would certainly be preferable to include railroad workers and firms, but this sector probably has few new firm formations, and those that do occur are primarily the result of entrepreneurs responding to right-of-way sales by larger railroads, so this is not a very serious exclusion.

The new businesses appearing in the BITS data identify new firms from their first filing of payroll tax withholding statements with the Internal Revenue Service. Because U.S. tax regulations consider owner/operators of corporations to be employees of the corporate entity, an incorporated self-employed person without additional employees will be considered a new employer-firm even when only the owner/operator is working there, while unincorporated firms will not be included until an additional person is hired and paid. This leads to some inconsistency in the distinction between zero, one, and two-employees businesses (of which there are large numbers), because the numbers of workers counted vary with the legal form of these businesses. Furthermore, because partners, owners of unincorporated businesses, and unpaid family members are not considered to be employees, and their take-home profits are not included in payroll, both the number of workers and their compensation is likely to be understated for non-incorporated firms. The microdata underlying the BITS includes a variable indicating legal form of business, but their code for 'incorporated' is also the code for 'unknown,' so no reliable data are available.

Because these data are limited to "employer-businesses," the formation of a new business can be identified only when the new business hires an employee and begins filing taxes for that employee. Similarly, the business will appear to close in any year that it lays off all its employees for an entire calendar year, because the BITS will not have data for any year with no positive payroll, even though the proprietor may continue the business activity alone. Therefore the closure of a business can only be tentatively identified in any year. Most businesses that remain without employees for two years, after having had employees in a prior year, are permanently closed. Nevertheless, a small fraction of the recorded new firm formations are actually reactivations of older firms that have not had any employees for over two years (nine quarters), so they are treated as new when they re-enter the SSEL.

As discussed above, the data underlying the BITS are quite comprehensive in their coverage of employment of firms, although some employment must be estimated from the reported first quarter payroll for those that fail to report the actual March 12 employment. For the smaller multi-location firms there are some weaknesses in the allocation of that firm-level employment to each of their component establishments. Firms with less that 250 employees are not surveyed every year to update their list of establishments and their distribution of employees among them, but are generally surveyed at least once between Economic Censuses. Firms with less than 10 employees are only asked about multiple locations during the Economic Censuses that take place every 5 years. New firms with less than 250 employees are assumed to be single-unit firms, so the proportion of new firms that have multiple locations is under-stated. However, this is not a central issue for entrepreneurship research or for geography.

The aggregate establishment and employments totals from the BITS are slightly lower that those from CBP, because the BITS has consolidated records for establishments that underwent mid-year reorganizations, while such establishments may appear in the microdata underlying the CBP under each of their identities, and therefore get double-counted. This would only be a problem when dealing with a small sector in a small region that had a big double-counted reorganization, which might lead to significant errors if CBP (on local employment or numbers of establishments) data were used to standardize BITS data, or to calculate rates of change.

Industry coding is often incomplete in the first year that a business appears, and a more detailed industry code may appear in subsequent years. In addition, the Census Bureau tends to place special emphasis on accuracy of industry codes in the year prior to Economic Censuses (which are in years ending with 2 or 7), because the censuses use industry-specific forms, so a higher proportion of records have changes in industry codes in those years, and there are fewer changes in the Economic Census year and the subsequent year. The BITS has fewer missing industry codes than CBP because the SUSB processing checks for updated codes from the subsequent year's SSEL and uses them to replace any missing or vague codes whenever possible.

All of the microdata in the BITS have been subjected to many layers of editing before being extracted into the BITS, so any anomalies that appear in the data for a given establishment are probably due to either lack of information in a particular year, or to actual eccentricities of the businesses' history. Just as we have found that the rates of new firm formation and closure are much higher than was realized several decades ago, more comprehensive microdata allowing the tracking of changes in establishments have shown that many businesses change their industry classification over time, often to completely different industry divisions. Many businesses also move across county lines, and even across state boundaries, often to non-adjacent states. Therefore the analyst must consider carefully which year to use for classification of industry or location, realizing that there may be differences across years during the normal life of the establishment.

One must carefully choose the most appropriate year for selection of data for classification of establishments or firms, in accord with the theory underlying the analysis, and with the practicalities of the database. Thus, for example, although industry coding tends to get more detailed and more accurate as an establishment stays in the SSEL, for coding the industry of new firm formations we chose to rely on the industry that was first reported, and only use later reports if the later industry code was a more-detailed variation of a less-detailed original code. This avoids the use of codes representing new industries that some businesses may shift into as they mature.

Although all establishments with data for a given year will have positive annual payroll for that year, nearly 10% are likely to have no employment. This may be because they had not yet hired anyone by the March 12 payroll period that is reported, or had closed before then, or are seasonal and do not operate in March, or had temporarily laid off everyone in March. The establishments without employees should be excluded to get counts of establishments active at a point in time. If they are included the count will represent all establishments that were in business at any point during the year.

The employment data are often estimates (especially for multi-unit establishments) based on reported payroll and average payroll per employee for similar establishments that did report. In these cases the resulting year-to-year comparisons are imperfect, and they probably tend to understate the actual employment changes.

Changes in employment may also be extreme, such as when a huge establishment sends most of its employees to work for an employment-leasing firm that handles all of its former personnel issues, or a new hospital staffs up with 3,000 employees in its opening year. Apparent

employment changes due to changes in business structure may be actual (when a firm consolidates two previously separate establishments), or misleading (as previously consolidated reporting for multiple establishments gets broken out more accurately, causing an apparent fall in the employment of the establishment where it was originally consolidated).

Mergers, acquisitions, and divestitures may cause large discontinuities in the characteristics of establishments and the firms they belong to. When a small single-location firm is acquired by a larger one, that establishment changes from single-unit to multi-unit, and its firm-size class is likely to change radically. Its own employment may also change, as some employees are laid off or reassigned to another location owned by the acquiring firm, and this may result in a reclassification of industry for the remaining portion of its business. When a new firm is created as a joint venture between two existing firms, it may appear as a huge new firm with a small new primary (headquarters) establishment and many large old secondary locations.

Finally, the BITS database is only available from 1989. The SUSB data commence with 1988, but the system for linking establishment data across time could not construct completely consistent linkages for the 1988 data. However, the CBP data have had consistent coverage since 1988, with only minor process-related variations around the Economic Census years (primarily more industry changes in the prior year, and more new multi-unit establishments and changes to mulit-unit status in the Census year). As with any large database, the BITS inevitably has some imperfections in its tracking of establishments that are changing their ownership or legal form, especially when changing between single-unit firms and mult-unit firms. Some linkages may be missed, causing false deaths and births, and some false linkages may occur between establishments that are not the same.

In spite of all these weaknesses, the BITS provides far more accurate and comprehensive data on U.S. private sector businesses than any other source. With few exceptions, in comparison with the rest of the world the BITS data offer broader coverage of the national and local economies, with more complete linkages between business firms and their owned establishments, and with more comprehensive tracking of establishments across ownership and legal changes.

VI. CONCLUSION

The BITS database constructed for the Office of Advocacy of the U.S. Small Business Administration is uniquely suitable for testing new approaches to explaining regional differences in economic growth rates. Recent theories of economic growth view local externalities, as opposed to scale economies, as the primary engine in generating growth in cities with their closely integrated surrounding counties (Labor Market Areas). While scale economies operate at the plant level, externalities operate at the level of the firm, primarily through entrepreneurial activity.

Using the BITS data we examined the impact of these externalities on regional employment growth from an entrepreneurial perspective by examining the relationship of local economic growth to local entrepreneurial activity. Since higher rates of entrepreneurial activity in an industry sector and region imply lower barriers to birth and greater local competition, this analysis can also be interpreted as an investigation of the impact of local competition on local economic growth. We found that higher rates of entrepreneurial activity were strongly associated with faster growth of local economies.

Our research leads to many more questions about the relationships between characteristics of economic areas, their rates of new firm formation, and their rates of economic growth. Further analysis of the BITS data, especially some of the more recent years of data, could reveal much more about the processes underlying the evolution and growth of the American economy, and how both new and small businesses contribute to these processes.

VII. REFERENCES

Acs, Z, and C. Armington, 2004a, "The Impact of Geographic Differences in Human Capital on Service Firm Formation Rates," *Journal of Urban Economics*, in press.

Acs, Z., and C. Armington, 2004b, "Employment Growth and Entrepreneurial Activity," *Regional Studies,* in press.

Acs, Z., and C. Armington, 2003a, "The Geographic Concentration of New Firm Formation and Human Capital," CES-WP- 03-05.

Acs, Z., and C. Armington, 2003b, " Endogenous Growth and Entrepreneurial Activity in Cities," CES-WP- 03-02.

Acs, Z., and C. Armington, 2000, Differences in Job Growth and Persistence in Services and Manufacturing," CES-WP-00-04 March .

Acs, Z., and C. Armington, 1998, "Longitudinal Establishment and Enterprise Microdata (LEEM) Documentation," CES 98-09.

Acs, Z., C. Armington, and A. Robb, 1999, "Measures of Job Flow Dynamics in the U.S.," CES-WP 99-01-January 1999.

Acs, Z. and D. Audretsch, 1990, *Innovation and Small Firms*, Cambridge: The MIT Press.

Acs, Z, D. Audretsch, P. Braunerhjelm, and B. Carlsson, 2003, "The Missing Link: The Knowledge Filter, Entrepreneurship and Endogenous Growth, Stockholm."

Acs, Zoltan, F. FitzRoy, and I. Smith. 2002. High Technology Employment and R&D in Cities: Heterogeneity vs. Specialization. *Annals of Regional Science* 36: 373-386.

Acs, Z., and A. Varga, 2002, "Geography, Endogenous Growth and Innovation," *International Regional Science Review*, 25(1), 132–148.

Acs, Z., and A. Varga, 2004, "Entrepreneurship, Agglomeration and Technological Change," *Small Business Economics*, in press.

Agion, P., and P. Howitt, 1992, "A Model of Growth through Creative Destruction," Econometrica, 60: 323-51.

Anselin, L., A. Varga, and Z. Acs, 1997, "Local geographic spillovers between university research and high technology innovation," *Journal of Urban Economics* 42: 422-448.

Armington, C. and Z. Acs, 2004, "Job Creation and Persistence in Services and Manufacturing," *Journal of Evolutionary Economics*, 14: 309-325.

Armington, C, and Z. Acs, 2002 "The Determination of Regional Variation in New Firm Formation," *Regional Studies*, 36: 33-45.

Armington, C., and A. Robb, 2002, "Mergers and Acquisitions in the United States: 1990–1994," CES 98–15.

Arrow, K. 1962. "The Economic Implication of Learning by Doing," *Review of Economics and Statistics* 80: 155–173.

Audretsch, D. and M. Keilbach, 2004, "Entrepreneurship Capital and Economic Performance," *Regional Studies,* in press.

Axtell, R. L., 2001, “Zipf Distribution of U.S. Firm Sizes,” *Science*, 293, 1818–1820.

Bartik, T.J. 1989. Small Business Start-ups in the U.S.: Estimates of the Effects of Characteristics of States,” Southern Economic Journal, 55: 1004–18.

Boden, R., 2000, “Establishment Employment Changes and Survival, 1992–1996,” Office of Advocacy, U.S. Small Business Administration, Washington D.C.

Braunerhjelm, P., and B. Borgman, “Geography, Entrepreneurship and Regional Growth,” *Regional Studies*, in press.

Ciccone, C. and R.E. Hall, 1996, “Productivity and the Density of Economic Activity,” *American Economic Review* 86:54-70.

Davis, S.J., J. Haltiwanger, and S. Schuh, 1996, *Job Creation and Destruction*, MIT Press, Cambridge MA.

Evans, D. and Leighton, L. S., 1990. Small Business Formation by Unemployed and Employed Workers. *Small Business Economics* 2: 319-330.

Evans, D. and B. Jovanovic, 1989, “Estimates of a model of entrepreneurial choice under liquidity constraints,” *Journal of Political Economy* 95: 657-674.

Feldman, M.P., and D.B. Audretsch, 1999, Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition. *European Economic Review* 43: 409-429.

Geroski, P. 1995. What do We Know About Entry? *International Journal of Industrial Organization* 13:421-441.

Glaeser, E.L., H.D. Kallal, J.A. Scheinkman and A. Shleifer, 1992, “Growth in Cities,” *Journal of Political Economy* 100:1126–1152.

Glaeser, E.L., J.A. Scheinkman and A. Shleifer, 1995, “ Economic Growth in a Cross- section of Cities,” *Journal of Monetary Economics* 36: 117–143.

Hannon M., and Freeman, J., 1989, *Organizational Ecology*, Cambridge: Harvard University Press.

Hayek, J., 1945, “The Use of Knowledge in Society,” *American Economic Review* 35: 520–530.

Headd, B., 2001, “Business success: Factors Leading to Surviving and Closing Successfully,” CES-WP-01-01.

Keeble, D. and S. Walker, 1994, “New firms, Small firms and Dead Firms: Spatial Patterns and Determinants in the United Kingdom,” *Regional Studies* 28: 411–427.

Kirchhoff, B.A., C. Armington, I. Hasan, S. Newbert, 2002, The Influence of R&D Expenditures on New Firm Formation and Economic Growth, Office of Advocacy, U.S. Small Business Administration, Research Summary #222.

Krugman, P., 1991, “Increasing Returns and Economic Geography,” *Journal of Political Economy* 99:483–99.

Jaffe, A., 1989, “The Real Effects of Academic Research,” *American Economic Review* 79(5): 957–970.

Jovanovic, B. and R. Rob, 1989, “The Growth and Diffusion of Knowledge,” *Review of Economic Studies* 56**:** 569–582.

Knight, F., 1921, *Risk, Uncertainty and Profit*, Houghton Mifflin, Boston.

Krugman, P. 1991, “Increasing Returns and Economic Geography.” *Journal of Political Economy* 99: 483–99.

Krugman, P., 1998, “Space: The final frontier,” *Journal of Economic Perspectives* 12, 161–174.

Lee, S. Y., R. Florida, Z. Acs, 2004, “Creativity and Entrepreneurship,” *Regional Studies*, in press.

Lucas, R., 1988, “On the Mechanics of Economic Development,” *Journal of Monetary Economics*, 22, 3–39.

Michelacci, C., 2003, “Low Returns in R&D due to the Lack of Entrepreneurial Skills,” *The Economic Journal*, 113 (January), 207–225.

Plummer, L A., and Z. Acs, 2004, “Penetrating the Knowledge Filter,” *The Annals of Regional Science*, in press.

Rauch, J.E., 1993, “Productivity Gains from Geographic Concentration of Human Capital:

Evidence form the Cities,” *Journal of Urban Economics* 34:380–400.

Reynolds, P. D., B. Miller and W. R. Maki, 1995, “Explaining Regional Variation in Business Births and Deaths: U.S. 1976–88,” *Small Business Economics* 7(4): 387–407.

Reynolds, P. (1994) “Autonomous firm dynamics and economic growth in the United States, 1986–1990,” *Regional Studies* 28(4): 429–442.

Robb, A., 1999, “New Data for Dynamic Analysis: The Longitudinal Establishment and Enterprise Microdata (LEEM) File,” CES 99–18.

Romer, P., 1990, “Endogenous Technical Change,” *Journal of Political Economy*, 98: 71–102.

Romer, P., 1986, “Increasing Returns and Economic Growth,” *American Economic Review*, 94, 1002–1037.

Schumpeter, J., 1911 (1934), *The Theory of Economic Development*, Harvard University Press, Cambridge, MA.

Shane, S. and J. Eckhardt, 2003, “The Individual-Opportunity Nexus,” in Z, Acs and D. Audretsch, eds. *Handbook of Entrepreneurship Research,* Boston: Kluwer.

Simon, C. and C. Nardinelli, 2002, “Human Capital and the Rise of American Cities, 1900–1990,” *Regional Science and Urban Economics* 32:59–96.

Storey, D.J. 1991. “The Birth of New Firms—Does Unemployment Matter?” A Review of the Evidence,” *Small Business Economics* 3:167–178.

Tolbert, C.M. and M. Sizer, (1996) “U.S. Commuting Zones and Labor Market Areas: a 1990 Update,” Rural Economy Division, Economic Research Service, U.S. Department of Agriculture, Staff Paper No. AGES-9614.

U.S. Small Business Administration, 1988a, *Uses and Limitations of USEEM/USELM Data*, Office of Advocacy, Washington, D.C.

U.S. Small Business Administration, 1988b, *Handbook of Small Business Data*, Office of Advocacy, Washington, D.C.

U.S. Small Business Administration. 1999, *Statistics of U.S. Business—Microdata and Tables of SBA/Census Data*, Office of Advocacy, Washington, D.C..

Van Stel, A. J., D. Storey, 2004, “The Link Between Firm Births and Job Creation: Is there a Upas tree effect?,” *Regional Studies*, in press.

Villalonga, B., 2001, "Diversification Discount or Premium? New Evidence from BITS Establishment-Level Data," CES-WP-01-13 December 2001.

White, L., 2002, "Trends in Aggregate Concentration in the United States," *The Journal of Economic Perspectives*, 16, 137–160.

Williamson, O., 1985, *The Economic Institutions of Capitalism*, New York: The Free Press.

Winter, S., 1984, "Schumpeterian Competition in Alternative Technological Regimes," *Journal of Economic Behavior and Organization*, 5, 297–320.

VIII. ENDNOTES

[1] The research reported here was initiated and supported by the Kauffman Center for Entrepreneurial Leadership at the Ewing Marion Kauffman Foundation, as the first step of a larger project to analyze the causes of regional differences in new firm formation rates in the United States. This initial research was carried out at CES under the title, "U.S. Geographical Diversity in Business Entry Rates" (Armington and Acs, 2002). Further research was funded by the NSF under Grant # SES-0080316, and was carried out at CES under the title, "Evaluation of New Service Firm Entries in the Standard Statistical Establishment List (SSEL) and Analysis of Regional Differences in their Entry Rates," described in working papers CES 03-05 (2003) and CES 03-01 (2003). The research examines the regional variation in entrepreneurship (Armington and Acs, 2002), the impact of human capital on entrepreneurship (Acs and Armington, 2004) and the relationship between employment growth and entrepreneurship, (Acs and Armington, 2004).

[2] For documentation of the SUSB files, see U.S. Small Business Administration (1999).

[3] For example, if a single-unit retail firm/establishment with 10 employees opens 2 additional branch stores with another 10 employees in each, the original establishment would appear to have 30 employees until it was surveyed. The more accurate reporting resulting from the survey would lead to its employment being reduced from 30 to 10, and the 2 new establishments being listed as new formations with 10 employees each.

[4] Taking the match of 1993 to 1992 as a typical example, 5.56 million records matched on CFN, another 32,000 on PPN, and 3,000 on EIN. The remaining unmatched single-unit records were then grouped by ZIP code, and another 19,000 between years and 24,000 within 1993 were matched on business name, and another 11,000 across years and 13,000 within 1993 were matched on industry (3-digit SIC) and street number.

[5] See also Knight (1921), Hannan & Freeman (1989), Acs & Audretsch (1990), Winter (1984), and Williamson (1985).

[6] In fact, birth rates were calculated for each annual period from 1990 through 1998, but these were found to be quite consistent in their rank ordering across LMAs, so the averages of several recent years was used for most of this analysis. Using period averages serves both to smooth out irregularities and to minimize the possibility of disclosure problems with very small numbers of annual births for the smaller LMAs and subsectors.

[7] Annually, there were less than 150 such large apparent births of single-unit firms, with an average of about 1,500 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. However, examination of the new firms with 100–499 employees in their first year showed that most seemed to be 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 employment impact of startups, there is no danger of the bulk of the data on smaller startups being swamped by that 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.

[8] 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 more credible for the first year.

[9] There is a small number (10,000 to 16,000) of new firms each year for which no industry code is ever available. Most of these are small and short-lived. These have been added to the Local market category, which is, by far, the largest of our sectors.

[10] Summary of research reported more fully in Armington and Acs (2002).

[11] These can be calculated from the ordinary coefficients, but it is more illuminating to view them as being estimated from standardized variables. In this case, rather than using the levels, ratios and percents of the exogenous variables discusses, we transform each variable by subtracting its mean value (calculated from all 394 LMA values) and then divide this adjusted value by the standard deviation of all 394 values. Each of these transformed variables has a mean of zero and a standard deviation of one, and each value represents the deviation of that particular LMA from the mean of that variable. Since the 394 LMAs constitute the universe at a point in time (rather than a sample of areas), it is apparent that the resulting standardized beta coefficients can be interpreted quite simply as measures of the impact of one standard deviation of the independent variable on the standardized dependent variable. For example, using standardized variables, if we estimate that x = .1y + .5z , then we can say that each standard deviation in the value of y is associated with 0.1 of a standard deviation of x, and each standard deviation of z is associated with half of a standard deviation of x. Obviously, it follows that x is five times more sensitive to z than to y.

[12] Summary of research reported more fully in Acs and Armington (2003a).

[13] Although we have annual firm formation data for 1990 through 1999, we have chosen not to use pooled cross-section time series regressions, because most of the independent variables describing the characteristics of the LMAs change very little over time, and the errors from omitted variables will be nearly identical for each LMA from year to year, so the diagnostic statistics from such an analysis would be very misleading.

[14] These numbers have increased considerably since then, but more recent data on educational attainment from the 2000 Census of Population had not yet been released at the county level, which is needed to construct the LMA level data.

[15] Note that when estimated in separate equations for 1996-98 the coefficient for College degree falls to .10 and that for high school dropout falls to .12, while other coefficients remain substantially the same.

[16] We originally hoped to base this classification on the BLS occupational distribution data for each (three-digit) industry, but we found that many activities requiring academic skills or advanced training for leadership positions, in fact had occupational distributions very heavily weighted toward semi-skilled and unskilled workers. Hospitals and hotels were extreme examples of this contrast between educational requirements for workers and those for the individual responsible for starting the business.

[17] The subsector classifications for each 4-digit SIC can be found in the Appendix of the CES Discussion Paper 03-02, where they are ordered by SIC code within each subsector. Data on the number of establishments and employees in each 4-digit SIC in 1995, and their net changes to 1998, as well as the total number of new firm formations during 1996 through 1998 per hundred (1995) establishments, are provided for each entry in this table.

[18] Summary of research reported more fully in Acs and Armington (2003b).

[19] When the new primary location of a multi-unit firm has less than a third of the total employment of the firm, it is not counted as a birth. Such relatively small new headquarters establishments are usually created to manage a new firm created as the result of a merger or joint venture, involving the restructuring of older firms.

[20] The number of firm births by LMA and sector in 1994 was not easily available, but had been shown consistent with the previous and subsequent years for more aggregated annual birth data.

[21] In an earlier paper (Armington and Acs, 2002), we regressed agglomeration effects on the firm birth rate. The results were positive, suggesting that greater density leads to more new firm formation. This suggests that higher density leads to greater creativity and spillovers (Lucas, 1989). However, it appears that growth is promoted by lower density.