Research universities and regional high-tech firm start-ups and exit

De Silva, Dakshina G. and McComb, Robert P.
Department of Economics, Texas Tech University

26. January 2009

Research Universities and Regional High-Tech Firm Start-ups and Exit

Dakshina G. De Silva and Robert P. McComb*[a]

January 26, 2009

Abstract:

If localized knowledge spillovers are present in the university setting, higher rates of both start-ups and/or survival than in the broader economy would be observed in areas that are geographically proximate to the university. Using a fully-disclosed Quarterly Census of Employment and Wages for Texas for the years 1999:3-2006:2, this paper analyzes start-ups and exit rates for high-tech firms in Texas. We find that there is evidence that the presence of a research institution will affect the likelihood of technology start-ups. However, results suggest that geographic proximity to knowledge centers does not reduce hazard rates.

Key Words: Entry and Survival, R & D, Regional, Urban, and Rural Analyses.

JEL Classifications: R12, R53, O18

---

* Department of Economics, Texas Tech University, Lubbock, TX 79032.
a Corresponding author: Email: robert.mccomb@ttu.edu, Tele: (806) 742 2466-240, Fax (806) 742 1134.
We would like to thank Hojin Jung and Anita Schiller for their skillful research assistance and Texas Workforce Commission for providing the data. We are also indebted to two anonymous referees, Georgia Kosmopoulou, and IIOC 2007 Conference participants for valuable comments and suggestions.

I. INTRODUCTION

During the decade of the 1990s, a significant level of firm start-up activity was observed in the newly emerging high-technology industries. This activity tended to be concentrated in relatively few locations such as Silicon Valley in California, the Route 128 corridor in Boston, and the I-35 corridor in Texas. Since the regional employment dynamism and relatively high incomes associated with these new technology firms were widely coveted by regional policy makers, regional economic development interests focused on initiating or attracting high-tech industrial "clusters" by looking to exploit the presence of correlates.[i] Chief among these correlates has been the presence of a research university or institution, or a broader research complex.

In this paper, we seek to estimate the effect of federally-funded R&D in universities and related research complexes in Texas on the likelihood of high-technology firm entry and survival. By restricting the analysis to Texas, we control for state-specific conditions across counties that influence the variables of interest and gain fully disclosed access to a highly detailed industry data set at the 6-digit level of the North American Industrial Classification System (NAICS).

Previous researchers have considered the question of the effect of university research on the regional economy. We are, however, unaware of any previously published paper that analyzes hazard rates of firms in terms of geographic proximity to knowledge generators.

As Nelson (1986), Jaffe (1989), Acs et al (1992, 1994), Acs et al (2002) and Fischer and Varga (2003) point out, it is quite plausible that the presence of a research university can make locally specific contributions to the level of commercial innovation in its region.[ii] The university provides geographically specific access to resources such as libraries, faculty, and a ready pool of graduates at all levels. Research universities and institutions conduct basic research, i.e., create knowledge, with the purpose of diffusing the knowledge they create. New knowledge that spills over most readily into the locality should result in localized private sector innovation. Moreover, universities increasingly seek means by which to facilitate faculty start-ups and to enhance access to university resources to support regional entrepreneurs.[iii] While universities can be sources of direct spin-offs in the form of start-ups, this impact seems to be moderate and relatively recent. The Association of University Technology Managers reports that 462 new high-technology companies based on academic discovery were formed in 2004 by 191 institutions, up 23.5% from 2003.[iv] This only represents an average of somewhat more than two start-ups per institution. In Texas alone in 2004, there were 787 start-ups in the high-technology activities.

If there is a geographic component to diffusion of knowledge, rapid innovation of new knowledge will enhance the economic value of geographic proximity to the knowledge production location. Moreover, given the publicly funded nature of university research, spillovers may be relatively more available from universities than private sector firms conducting similar R&D. Jaffe (1989) and Jaffe et al (1993) find evidence of localized knowledge spillovers from universities. In particular, they find that the presence of a university positively affects the local or regional level of patent activity. Anselin et al (2000) find evidence that university spillovers are specific to certain industries. For example, their results suggest the strong presence of spillovers in the case of electronics but not in drugs and pharmaceuticals (at the 2-digit SIC level). Mansfield (1995) also finds evidence that the level of university R&D expenditures and quality of relevant faculty are important to industrial innovation in technology industries (at the statewide-level). He also recognizes the importance of the nearby presence of other firms in the same industry.

Zucker et al (1998) stress the importance of basic research and the growth and location of human capital to the location of biotech start-ups. They place an emphasis on the location of highly productive university faculty in the life sciences. Looking at data for 183 functional economic areas (defined by the BEA) for the period 1976-89, they find that the number of both highly productive faculty and top-quality research institutions had a positive and significant effect on the geographical distribution of the stock of bio-tech firms in 1990. They also find that the number of faculty "stars" had a positive, but diminishing effect on the number of bio-tech start-ups over the period. They underscore the importance of start-ups to the growth and geographical distribution of the biotech sector. As they note, the biotech industry went from being practically non-existent in 1975 to over 700 firms by 1990. In a similar way, the rapid pace of innovation in the new high-technology activities that emerged in the 1990s appears to have been expressed in high rates of start-ups and exits.

While we might posit that the presence of local spillover effects from university research should contribute to higher levels of technology start-up activity and lower rates of exit in more proximate geographic regions, we can not necessarily disallow other possibilities for the localized influence of the universities and research centers. It may be that the presence of a large university is a driver for the social and cultural climate that creative individuals require. In such a case, it is not just the technical knowledge that is created and locally diffused by the university, but also the aura of the university that matters in the attraction or retention of creative technology entrepreneurs. More generally, the higher the quality of the social, recreational, and cultural environment, the more attractive the area should be to entrepreneurs surveying locales in which to start their firm.

By the same token, we cannot eliminate the possibility that the relationship between the size of the research institution and the quality of the pool of recently trained individuals is positive. If these graduates are seeking a means to earn a living within their alma mater's community, then perhaps better trained graduates are better prepared to initiate a start-up. But this represents an additional channel for localized knowledge spillover.

Although the presence of university research activities may be a necessary condition for the presence of a significant cluster of high-tech private sector activity, it clearly is not sufficient. Examples of relatively undeveloped regional economies that host notable research universities are not difficult to find. Indeed, while much of the research to date has found a positive effect of university R&D on growth of high-technology industrial activity, the estimated effect is generally rather modest. There clearly are other important determinants.

In an approach similar to our paper, Woodward, Figueredo, and Guimarães (2006) consider the effect of university R&D in science and engineering on the appearance of new high tech establishments across all contiguous counties in the U.S. from 1997 through 2000. Using Census data at the 3-digit SIC level to identify high tech entrants, they analyzed a set of 31 "R&D intensive" industries. Controlling for factor costs, urbanization and localization economies, and cultural and natural amenities, they find a significant effect of university R&D on the probability of a firm locating within close proximity of the research university. However, the effect is slight. They conclude that a $1 million increase in university R&D increases the probability of a new establishment by less than 0.1 per cent.

Abramovsky, Harrison, and Simpson (2007) investigate the effect of university R&D on the location choices of R&D-performing firms in Great Britain across 111 postal zones. Using a quality index for university research departments, the authors construct a weighted-average index for all universities and aggregate measures of the departmental-level research quality in ten research fields in each of the regions. Looking at the geographic distribution of both existing establishments and the number of new entrants of R&D-performing firms over the period 2001-2003 for six product groups, the authors find little empirical support for the proposition that the number or quality of research universities had a positive effect on the location choices of R&D-performing firms. Oddly enough, where the universities appear to have some relevance to location decisions, they are lower tier, not top tier, research institutions. Agglomeration effects, industry-specific localization economies and workforce educational characteristics appear to play important roles.

The results of our analysis suggest there is persuasive evidence that the presence of a research institution and the size of its research enterprise affect the probabilities of technology start-ups in Texas counties. On the other hand, there is little support for the hypothesis that geographic proximity to knowledge centers reduces hazard rates. This latter finding runs counter to the proposition that localized knowledge spillovers are present. Section II describes the data set. Section III reports the results of the empirical analysis; and Section IV summarizes the main findings of the study.

II. DATA

We obtained firm-level data for Texas from the Quarterly Census of Employment and Wages (QCEW) from the Texas Workforce Commission. This data set provides firm-specific monthly employment and quarterly total wages reported by establishment as required under the Texas unemployment insurance program. Each record includes the specific location (address) of the establishment, business start-up date (the date on which UI liability begins), and the relevant six-digit NAICS code. Furthermore, separate establishments (branches or franchises) of the same firm are separately identified and reported in separate records. This panel data set is comprised of quarterly observations for each firm from Q3:1999 through Q2:2006. Each record includes each firm's unique Employer Identification Number (EIN). It should be pointed out that the authors obtained these data under an agreement of confidentiality and disclosure of the actual data is subject to certain restrictions.

The focus of this analysis is on the technology or knowledge-intensive industries. Definition of the technology sector is relatively easy in the abstract, but difficult in reality. There are industries that commercialize new or adapted technologies and industries that are relatively more dependent on applications of technology to remain competitive. This paper adopts the definition of technology industries as developed by

Paytas and Berglund (2004). Their classification identifies technology employing industries as those industries in which employment in technology occupations is at least 3 times the national average.[vii] Primary technology generating industries are defined as those industries in which R&D expenditures per employee and the proportion of scientists and engineers in the workforce exceed the national average. Secondary technology generating industries are those that only meet one of the two criteria.

While BLS occupational data are only aggregated at the 4 and 5-digit NAICS, Paytas and Bergland translate these into 4 and 6-digit NAICS industry classes. This yields a list of primary and secondary technology generators, and a list of technology employers at 6-digit NAICS codes. We use these NAICS-6 codes to identify technology firms. In our data set, we have more than 17,000 technology firms and 900,000 total observations. From these, we identify the numbers of start-ups, incumbents, and exits (by NAICS code) for a given county for a given fiscal year. Figure 1 illustrates the concentration of these high-tech start-ups in the metropolitan areas of Texas along Interstate 35. One can also note a sprinkling of high-tech firms across the less urban areas of the state.

Figure 1 about here

Using the QCEW, annual county employment in each 6-digit NAICS code is computed as the average monthly employment level for the year. The year is defined as the four consecutive quarters beginning with the third quarter in each calendar year. To arrive at annual average income for each industry, we compute total average income within each 2 digit NAICS for the year.[viii] While it would be ideal to have a narrower definition of industry for purposes of average income, it is not practical since many counties have no employment, and thus no income, at the more detailed industry level. Working at the NAICS-2 provides a non-zero datum on factor cost for each county for every industry.

We are interested in the likelihoods of start-ups and firm survival given geographic proximity to knowledge centers. Hence, the number of start-ups and exits by NAICS code for a given county for a given year are the units of observation. There are only fifty industries being tracked. Four high-tech industries at 6-digit NAICS were omitted since there were no start-ups over the period of this analysis (see Table A1). We have then 76,200 observations over the six years 1999:3-2005:2.[ix]

The appearance of a new EIN is used to define market entry and disappearance of the EIN is treated as an exit during any given period. Exits, in this context, may not signal business failure since the initial objective of many technology entrepreneurs (and their venture capitalist backers) is to build a firm with the intent of selling the firm within a one to three year time frame to a larger, incumbent firm. In some cases, this sale will result in the disappearance of the original firm from the data set if the firm is merged into another establishment, re-launched as a new firm, or relocated outside the county. While this may complicate the analysis of the effect of university R&D on firm failure rates, there is no reason to suspect *a priori* that the proportions of firms that either fail or sell should vary systematically, *ceteris paribus*, by geography. Therefore, we assume that the variability across counties of firm failures is directly proportional to the variability of firm exits from the data set. This method to identify exits has been used elsewhere in the economics literature. For example, Baldwin and Gorecki (1991), Dunne et al, (1988, 1989a, 1989b, 2005) used this definition of exit in their analyses of firm entry and exit.

Table 1 about here

Table 1 provides basic summary statistics for start-ups and incumbents. On average, for a given six digit NAICS code, there are about .7 incumbents and .07 new firms per county per year. When considering the number of exits, on average, there are about .11 exits by incumbent firms per county per year and .063 by start-ups. There were no start-ups or incumbents in many counties during the period of this study. Looking only at counties in which high-tech firms are located, incumbents have on average about 10 more employees per month than start-ups. Also, monthly income for start-ups and incumbents differs significantly. Start-ups pay about $300 per month less than established firms. For MSA counties, the average number of incumbents is 2.1 and the average number of start-ups is .2. In the case of knowledge center counties, the average numbers of incumbents and start-ups are 12.1 and 1.2, respectively. Nevertheless, in both the MSA and knowledge center counties, the differences in numbers of employees and wages between start-ups and incumbents are similar to what we observe for all counties.[x]

We proxy the level of research activity within the knowledge centers by using total federal research awards by federal fiscal year to Texas universities and research institutions for science and engineering R&D. This represents the magnitude of potential knowledge spillovers from research universities and institutions. Data on university R&D expenditures were obtained from the National Science Foundation. These expenditures are available by recipient institution by granting agency for each year of our analysis.

Although the NSF provides research funding by institution that is identified by granting agency or departmental source, i.e., DoE, EPA, DoD, we aggregate total federal awards by geographically distinct institution, i.e., system campuses are scored geographically separately. There is one significant exception to this geographic separation; the reported totals for Texas A&M (the state land grant institution) are partially aggregated in "Texas A&M, all campuses", although the Texas A&M Health Sciences Center is reported separately. Without additional specifics, we attribute the total awards to Texas A&M, all campuses, to the main campus in College Station.

Total external R&D funding is calculated on a county-by-county basis by adding all awards to all universities within a county. Since our objective was to identify universities and research centers that actively conduct R&D, we define a knowledge center as a county receiving at least $10 million (in 1999 dollars) in federal R&D funding during any federal fiscal year between 1995 and 2003.[xi] This bar captures the great bulk of externally funded R&D in Texas universities and medical research institutions. It reduces the 254 counties in Texas to ten counties deemed to host a knowledge center. Table A-2 identifies the academic institutions in these counties. Harris County (Houston) has consistently received the largest amount of federal funding followed by Travis (Austin) and Dallas (Dallas) counties. Table 2 presents the total funding of these selected counties.

Table 2 about here

To capture the local workforce characteristics relevant to high-tech activities, we considered two alternative formulations. One alternative is the share of county population with bachelor's degree or higher as reported by the U.S. Census Bureau in the year 2000 decennial census. Because the Census Bureau reports these data for only 116 of the 254 Texas counties, i.e., counties that are included in either a metropolitan or micropolitan statistical area, we would have to use the percentage reported by the Census Bureau for Texas for all rural counties outside of metropolitan areas, or 13.8%, for the remaining 138 counties. This compares to 23.2% for the state as a whole. Since there would be no variation in this variable for the majority of counties over the period of the analysis, we prefer to use the share of county population between the ages of 20 and 44. These estimates are available annually (year 2000 is actual census data) for all counties from the Census Bureau's Annual Population Estimates. This variable serves as a proxy for the technological savvy of the workforce and assumes younger workers are more comfortable with rapidly evolving technologies. This should also reflect the relative labor force potential for high-tech entrepreneurial activity insofar as much of the high-tech boom was reputed to be driven by relatively young entrepreneurs.

As Woodward et al (2006) suggest, cultural and natural amenities are important to industrial attraction and skilled workforce retention. Since climatic conditions do not vary substantially across the state, we focus on local cultural and recreational amenities and dining/hospitality options as the relevant variable. To measure the relative local presence of these amenities, we compute the share of county employment in NAICS 71, Arts, Entertainment, and Recreation, and NAICS 72, Accommodation and Food Services, as reported in the QCEW data set. The NAICS 72 activities also reflect the scope of the locality's amenities for business travelers and informal business and social interaction.

To account for factor costs, we use the yearly median rural land price in each of seven land market regions in Texas for the counties comprising the region as reported by the Texas A&M Real Estate Center. As a second measure, we use average wages paid in the entering or surviving firm's industry at the 2-digit NAICS as described above. The county unemployment rate for each year, as reported by the Texas Workforce Commission, is also included to provide an indication of the overall economic conditions in the county.[xii]

In the next section, we empirically model the relationship between the levels of federally funded R&D and the rates of technology firm start-up and survival patterns of technology firms.

III. EMPIRICAL ANALYSIS

<u>Research and Development Expenditures and Entry</u>

An important feature of this analysis is the relatively high degree of both geographical and industrial detail that is utilized. As the size of the geographical unit of analysis increases, the ability of the model to incorporate and detect the effects of localized factors decreases. For example, local agglomeration economies, local labor force characteristics, and the availability of cultural and recreational amenities probably play an important role in industry attraction but are increasingly obscured as the geographic area and diversity broadens. By the same token, the greater the level of industry aggregation, the more likely it is that informative industry characteristics are lost. Within a 2 or 3-digit SIC or NAICS code, industries can be quite heterogeneous in terms of inputs and relevant output markets. The geographical area for this analysis is the county-level which, in Texas, is relatively small. Counties are mostly square (see Figure 1) and, on average, only 44 miles across the diagonal. The industrial data at NAICS-6 are also highly detailed.

We first consider the number of new start-ups in the technology industry in Texas as a function of distance from knowledge centers and the level of external funding at these knowledge centers. This level of funding is a proxy for the "size" of the knowledge center or, alternatively, for the amount of knowledge being produced. Since we are examining the number of start-ups, we employ a count data model, a fixed effects Poisson model using six-digit NAICS codes as a group variable. This allows us to control for unobservable heterogeneities at NAICS levels.

A Poisson model assumes equality between the mean and variance of the dependent variable. Specification tests for over-dispersion reject the Poisson as the appropriate distribution for the data in this study. In such cases, researchers have often turned to the negative binomial model which differs from the Poisson in that the assumed distribution for the dependent variable exhibits over-dispersion. Although the negative binomial may be the preferred approach in many empirical settings, the Poisson has an important property which, we believe, makes it preferable in this context. It can be shown that Poisson Quasi-Maximum Likelihood Estimations (PQML) are consistent under quite weak assumptions (Gourieroux et al, 1984a,b).[xiii] In fact, the data do not have to come from a Poisson process at all, and may be either under or over-dispersed. Essentially, all that is required for consistency is that the conditional mean function is properly specified.[xiv]

Since we believe this condition is met, we use the fixed effects PQML estimator to produce our baseline results.[xv] In this case, the estimated coefficients are identical to fixed effects Poisson regressions but the standard errors are adjusted for over-dispersion. We report robust standard errors clustered by NAICS codes. Our dependent variable is the number of start-ups (y) for a given county (i) for a given year (t) by six-digit NAICS code (α). The basic model is as follows.

(1) $f(y_{it}|\mathbf{X}_{it}, \alpha_i, \boldsymbol{\beta}) = \prod_{t=1}^{T} exp[-\alpha_i \exp(\mathbf{X'}_{it}\, \boldsymbol{\beta})]\, [-\alpha_i \exp(\mathbf{X'}_{it}\, \boldsymbol{\beta})]^{y_{it}}/y_{it}!$

Then the conditional mean is:

(2) $E(y_{it}|\, \alpha_i, \mathbf{X}_{it}) = exp(\gamma_i + \mathbf{X'}_{it}\, \boldsymbol{\beta})$

where $\gamma_i = \ln\, (\alpha_i)$.

The independent variables **X** can be classified into four main groups: x_1 represents knowledge spillovers, x_2 controls for county *i*'s (herein base county) characteristics for a given year for a given industry, x_3 controls for the average characteristics of county *j* (contiguous neighbor of base county *j*) for a given year for a given industry, and x_4 are time dummies to identify fiscal years. Every county in Texas is treated as a base county in the analysis, and each is considered in relation to its distance from the relevant knowledge center county.

To account for decay of the spillover effect as distance to the knowledge center increases, we deflate total annual R&D funding in the knowledge center county by distance between *base county* and *knowledge center* county. This is done by computing the ratio of the knowledge center's total R&D funding to distance between the base county and the knowledge center county. A ratio is computed for all counties relative to all knowledge centers. The distance between counties and the knowledge center county is calculated as the distance in miles between the latitude and longitude of each county seat. Note that this construction results in a relatively rapid decay in federally funded R&D expenditures as the distance increases between base counties and the knowledge center county. In cases where the base county is also the knowledge center county, we assume the county is square, and take half the distance of the diagonal from opposite corners of the county. This avoids the problem of zero distance in the denominator of the spillover ratio for the knowledge center in these cases.

Each county is then assigned to the region of the knowledge center for which the R&D spending to distance ratio is greatest. The set of all Texas counties is thereby partitioned into 10 non-overlapping "spillover" regions in each period.

The question of the appropriate lag in R&D expenditures is also important. If knowledge spillovers from federally funded research are present, there will clearly be a time element involved. External funding for research in any given year supports that year's research. While it is clear that most scientific and engineering R&D projects are ongoing and take place over the course of several years, the research done in any given year will generally not be available for commercial purposes until at least the following year. Thus, whether it is ongoing research or newly established research initiatives that are being funded by federal support, the effect can reasonably be expected to occur in the following or several successive periods. Therefore, we lag the R&D funding by one to four years in order to address the time element in knowledge diffusion.

Table 3 provides a sense of the geographic size of the ten spillover regions in terms of the average minimum distance between the base county and its knowledge center county. It also includes the average normalized spillover size of each region by lagged period. Harris, Dallas, and Travis counties have the highest funding per mile.

Table 3 about here

When controlling for base county characteristics, we first use the past year's (t-1) log of average monthly employment by six-digit NAICS code for start-ups and incumbents at the county level.[xvii] These variables control for industry-specific localization economies due to the presence of specific labor resources, kindred firms, and a larger pool of potential entrepreneurs within the existing industry base. Experience working within an industry should enhance an entrepreneur's ability to enter that or a similar industry.

We also use year t-1's number of start-ups, and the number of exits by new firms and incumbents by six-digit NAICS code. This controls for the "entrepreneurial culture" in the locality. Entrepreneurial culture reflects such features as the presence of serial entrepreneurs, availability of venture capital, support networks for entrepreneurs, local attitudes toward business failure, local incentive programs to stimulate entrepreneurial start-ups such as business incubation programs, and aggressive technology transfer programs at the local knowledge center that facilitate local start-ups. Similarly, when controlling for neighboring counties, we use all the variables that control for base county characteristics.

In order to account for the presence of urbanization economies, we include a dummy variable that takes the value of one for a county that is part of a metropolitan statistical area (MSA) and zero if the county is not within an MSA. We use the 2004 MSA county list from the Texas State Data Center that includes several counties that were added to MSAs in 2003 based on results of the 2000 Census. The MSAs account for only 77 of the 254 Texas counties but contained about 86 percent of the state's population of nearly 21 million at the time of the 2000 Census.[xviii] Use of the qualitative MSA variable to capture agglomeration economies seems preferable to county manufacturing or population density measures since it provides a simple means to capture regional urbanization and labor draw for counties that may be largely suburban or partially overtaken by sprawl. As noted, we also include the county unemployment rate for the year.

Table 4 about here

The first set of regression results are reported in Table 5a. The model does a very good job in explaining the likelihood of technology firm start-up across Texas counties. In column 1, we lag the funding per mile by one year. In columns 2, 3, and 4 we lag the funding per mile by two, three and four years keeping the base county and neighboring county characteristics lagged by one year. In general, where the signs of the estimated coefficients are significant, they are also positive as would be expected if localized spillovers are important in explaining technology start-up activity. The probability of observing a start-up in the given county decreases as the distance from the knowledge center increases, all else equal.

When considering base county characteristics, the coefficients on the qualitative variables that identify counties as hosting a knowledge center and being part of an MSA are relatively large and significant. Total employment in previous start-ups in the same industry also appears to have a positive influence on entry. This suggests that the presence of an entrepreneurial culture is an important explanatory variable. Similarly, the higher the past employment in incumbent firms, the higher the rate of new entry into the market. This would be consistent with the view that employees of existing firms represent a pool of potential entrepreneurs (who prefer to start a business where they already live) and a local source of potential employees for an entrant. Both effects are also consistent with the view that the start-ups enjoy localization economies from the presence of other firms in the industry.

The estimated coefficient on the lagged number of exits by start-ups is negative. This suggests at least two interpretations. A higher number of exits from a given industry may increase the perception of start-up risk of failure and reduce local venture capital interest in that industry. Secondly, if the higher number of exits of recently launched firms in the county is a result of weak local entrepreneurial support networks, then the location will be less attractive to new potential entrants.

Higher average income in the industry in the county appears to increase the likelihood of observing a start-up. One reason may be that higher incomes in an industry provide a positive signal to potential entrepreneurs and attract entry into that industry. However, from the perspective of factor costs, theory would suggest a negative relationship between local labor cost and the likelihood of entry. It has, nevertheless, been generally observed that relative returns to high-skill labor grew substantially during this period of rapid growth in the technology sector. This variable may in fact be capturing the differences in occupational configurations within the NAICS-2 classification across counties. By the very definition of the high-tech sector at the NAICS-6, high-tech entrants require employees with higher technology skills than the broader class of firms within their NAICS-2. In such a case, if the specific skills they require are present in greater proportions in a given locality than elsewhere, the NAICS-2 income level in that locality should be higher. Given the importance of the presence of specific technology skills to this sector, a positive correlation between average wages and entry should perhaps not be surprising even though we control for lagged employment in the entrant's NAICS-6.

Despite the inclusion of dummy variables for both knowledge center (large research institution) and MSA (urban) county characteristics, both of which would correlate with relatively greater social and cultural opportunities (including spectator sports events), the variable measuring social, cultural and recreational amenities has a positive and quite significant estimate. The estimate of the coefficient of the proportion of population between the ages of 20 and 44 is also positive and quite significant, if not particularly large in magnitude. In terms of the broader economic conditions prevailing in the county, the softer the regional economy, as expressed in terms of unemployment rates, the less likely it is to observe a start-up.

Among the variables controlling for neighboring county characteristics, three have some statistical significance. It appears that the greater the number of lagged exits in a neighboring county within a given industry, the more likely it is that a start-up will occur. One possible explanation may be that higher exit rates of start-ups are a result of a relatively less supportive entrepreneurial environment. Thus, while exit rates within a county appear to have a negative influence on the probability of future start-ups in the same county, local entrepreneurs may seek a more supportive adjacent location within the region. The greater the value of the amenity variable in the neighboring county, the lower is the probability of start-up. Amenities, it appears, are also localized at the county level or below and valued by regionally mobile entrepreneurs. Lastly, the greater the proportion of the workforce between 20 and 44 in the neighboring county, the less likely it is to observe a high tech start-up in the base county. As before, this suggests that high-tech start-ups, at least at a regional level, gravitate toward localities with relatively younger workforces.

Table 5b about here

In order to check robustness of our chosen methodology, we re-estimate using the Dirichlet-Multinomial (DM) Regression method with NAICS-6 as group effects. This method was specified by Guimaraes and Lindrooth (2005, 2007) and also used by Woodward et al (2006) as an alternative method to estimate over-dispersed count data models controlling for group heterogeneities. They show that the DM model is a natural extension of McFadden's conditional logit model for grouped data and discuss its use for the case of count models. As can be seen in Table 5b, results from the DM estimation are qualitatively the same as the results from the PQML estimation.

Hazard Rates

Next, we consider hazard rates for all incumbents and start-ups between FY1999-2000 to FY2005-06. As noted, most previous studies have found that university research positively affects firms that are generally located near universities or research centers. If the effect is greater the closer a firm is to the knowledge center, then hazard rates would be expected to be lower *ceteris paribus* for technology firms located in geographical proximity to knowledge centers and to increase as proximity decreases.

We consider a standard non-parametric Kaplan-Meier approach to estimate the hazard functions for start-ups and incumbents. In our dataset, the average length of survival for start-ups is 27 months. Hence, we track survival up to 36 months. Table 6 provides survival probabilities for start-ups and incumbents. These results clearly show that the hazard rate for start-ups is substantially higher than the hazard rate for incumbents. After three years, the probability of survival for a start-up is only about one-third (36%) while that of incumbents is 92%.

Table 6 about here

Note that the Kaplan-Meier method is useful for comparing survival curves in two or more groups but it does not control for explanatory factors such as firm size. Cox proportional-hazards regression allows analysis of the effect of several risk factors on survival. In this case, we can examine the probability of exit controlling for spillover, MSA, size of the firm, average wage of the firm, and market conditions.[xix] As a measure of firm size, we use the average ratio of the number of employees in the given firm to the total number of employees in the corresponding six digit NAICS in the given county over the period of the firm's operation. The basic Cox proportional hazard model can be written as follows:

(3) $$h(t|\boldsymbol{z},\boldsymbol{\psi}) = h_0(t)\exp(\boldsymbol{z}'\boldsymbol{\psi})$$

where $h(t|\boldsymbol{z},\boldsymbol{\psi})$ is the conditional hazard rate and $h_0(t)$ is the unspecified baseline hazard function. The vector of covariates are denoted by $\boldsymbol{z}$ and $\boldsymbol{\psi}$ are the corresponding coefficients estimated by the Cox regressions. The predictors are the spillover variables, distance variables, establishment's average wage, relative employment, and market conditions as represented by the prime rate and the state monthly unemployment rate.[xx] The spillover effect is captured by three dummy variables constructed by dividing the range of the distribution of the average spillover that a firm experienced during its existence into four equal intervals and treating membership in the bottom quarter of the range as the omitted group. Distance from the knowledge center is captured by two alternative means. In columns 1 and 2 of Table 7, distance is measured in miles, as explained above, and in columns 3 and 4 it is measured using distance dummies.

Table 7 about here

Table 7 contains the first set of Cox proportional hazard estimates for start-ups and incumbents. As expected, we find that the hazard rate for startups is higher than for incumbents, all else equal. The results also indicate that the larger the firm, the longer is the expected survival period of the firm. This is consistent with the findings of Dunne et al (1989a and 1989b) that firm size matters. Looking at the spillover dummies and interaction terms, it appears that the spillover has no significant positive effect on the likelihood of survival for either entrants or incumbents. While the estimated coefficients on the distance variables suggest a weak negative effect on firm survival probability, we find it difficult to imagine that closer proximity to a research institution can increase the likelihood of failure. We do not therefore find this result to be very compelling and assume that another unobservable influence is at work. For example, the greater concentration of firms in these locales may result in a more continuous distribution of firms across risk classes.

As Audretsch and Mahmood (1995) found, the wage rate and prime rate are negatively correlated to the hazard rate while the unemployment rate is positively correlated. As they suggest, higher interest rates do not directly affect these firms since they do not rely on bank financing for external capital. Unlike Audretsch and Mahmood, we estimate the hazard model separately for interest rates and unemployment rates because these two variables are highly correlated (-0.9442).

The proportional hazards assumption assumes that the hazard ratio is proportional over time. A common method of evaluating the proportional hazards assumption is to plot the Kaplan-Meir (KM) observed survival cue with the Cox predicted curve and compare. When the two curves, actual and predicted, are close together, the proportional hazards assumption is not violated. From Figure 2, we can clearly see that the two curves are quite similar for both start-ups and incumbents. Note that Figure 2 is drawn after estimating the first column in Table 7. The alternative specification in the second column of Table 7 also shows that the proportional hazards assumption is not violated (figure not provided to save space).

Table 8 provides the distribution of new firms and incumbents by distance. Out of 7,713 high tech firms entering the market between July 1999 and June 2005, about 84% were located within less than a 50-mile radius from knowledge centers. On the other hand, only about 5.75% were beyond the 100-mile radius. Almost identical proportions are observed in incumbents' distribution as well. While suggestive, this also reflects population distribution.

Table 8 about here

Figure 2 about here

In Figure 3, we graph the survival patterns for start-ups and incumbents by spillover group. It is clear from this graph that there is a significant difference in survival rates between start-ups and incumbents. However, if either group is considered separately, distance does not have any apparent effect on the within-group survival rates. This is consistent with the results reported in Table 8.

Figure 3 about here

As discussed above, if knowledge spillovers represent positive external economies to technology firms located close to the location of the R&D activity, hazard rates would be expected to be lower, *ceteris paribus*, for technology firms located in geographical proximity to knowledge centers and to increase as proximity decreases. Figure 4 illustrates hazard rates for start-ups and incumbents by the minimum distance to knowledge centers. Again, from these graphs, it is clear that there is a significant difference in survival rates for start-ups and incumbents. However, as in the previous graphs, within groups, there is little apparent effect on hazard rates from distance to the knowledge center. For new start-up firms within the 50-mile radius, the survival rate is about 35% after 36 months compared to a rate above 90% for incumbents.

Figure 4 about here

Note that our current definition of start-ups limits the period of analysis. For example, we may observe a firm entering the market in May of 2005. But since our data set ends in June of 2005, the Kaplan-Meier Survival Function Estimates would treat this firm as existing only one month before exit, since all firms appear to terminate with the end of the data in June, 2005. By including these late entries, there would be a tendency to under-estimate the survival rates for new firms. To overcome this problem, we change our set of observations on start-ups to include only those firms that can be tracked for at least three years. We drop from the sample all firms that entered after 36 months before the end of the data, i.e., any firm that entered after June 2002. Hence, we treat as a start-up any firm that entered the market between July 1999 and June 2002 and trace their existence for three years from the date of start-up. To be consistent with the new firm definition, we treat an incumbent as a firm that entered the market before July 1996. Firms that entered the market between July 1996 and June 1999 will be dropped since they will not have already survived at least 36 months. Hazard Rates for all start-ups and incumbents under the alternative definitions are illustrated in Figure 8. To economize on space, we provide only the graphical results.

Even with our alternative definition our qualitative results do not change although the magnitude of survival of new firms has increased. This would be expected with our new definitions. This qualitative result is true for all above estimated hazard rates.

Figure 5 about here

IV. CONCLUSION

The intent in this study was to consider the localized influence of university R&D on the likelihood of firm start-up and survival. In the case of start-ups, our assumption is that, if a localized spillover is present, it will be expressed in a higher likelihood of observing new firms in the knowledge-intensive or high-tech industries in areas more proximate to the university. As suggested at the outset of this paper, the knowledge spillovers are most easily pictured for us in the form of start-ups based on intellectual properties that result either directly from the research or as variations or derivatives of that research. This would be the case if, for example, the research results and implications are diffused locally through informal networks before the research outcomes are made widely public.

While otherwise controlling for the qualitative effect of the presence of a research institution, we find evidence that both the size of the research enterprise and its relative proximity help to explain the likelihood of start-ups in a locality (county). This is consistent with the hypothesis that specific knowledge spillovers are present. As distance to the knowledge center increases, for a given level of university R&D, the likelihood of a start-up decreases. Computing marginal effects (for Column 1 in Table 5a), we estimate that a 1 percent increase in the distance from the knowledge center, all else equal, will result in a .1191 percent decrease in likelihood of observing a start-up in a given county. This also suggests a relatively sharp diminution in likelihood of observing a start-up as distance from the knowledge center increases, for any given level of R&D funding. This finding is consistent with previous research on the question (see, for example, Woodward et al, 2006). Other factors appear to be more important than the level of R&D funding. Specifically, other localized non-research elements embodied in the presence of the university and the area's metropolitan characteristics have a greater effect.

The question of the effect of university R&D in the case of firm survival is somewhat different from that of entry. A start-up firm based on intellectual property that devolves from or is motivated by university research may or may not benefit directly from continued proximity to the on-going university research after the firm is actually launched. One can easily imagine an entrepreneur enlightened by the university research who then conceives an innovative idea, starts a firm, and is off and running on his/her own. On the other hand, continued localized benefits in the form of technical knowledge resources (brown bags, libraries, and easily accessible faculty consultants), related on-going research, or a greater pool of skilled labor will extend competitive advantages to nearby firms. If so, firms located in closer proximity to the knowledge center should exhibit higher survival rates than more distant firms with reduced access to these economies.

We find weak evidence that hazard rates are positively influenced by proximity to the knowledge centers. The question we have posed is whether or not knowledge spillovers from university R&D reduce hazard rates. The results of this analysis, then, would indicate that beneficial spillovers are not present or that spillovers do not enhance the likelihood of firm survival regardless of distance from the knowledge center.

Hazard rates for start-up firms are, of course, significantly higher than for incumbent firms. Consistent with the findings of Dunne et al (1989), we find that larger firms are more likely to survive. Market conditions also appear to matter insofar as the unemployment rate has a positive and significant effect on hazard rates.

One further conclusion that might be drawn from these results is that if presence of a research university does not influence hazard rates, as it appears in this study, it can make a contribution to regional economic activity by inspiring higher rates of start-up activity. That is, higher rates of start-up activity at similar hazard rates will result in higher levels of technology industry employment in the regions around knowledge centers. Moreover, since the presence of technology firms is also a factor in explaining technology start-up activity, the growth in technology firm start-ups should accelerate relative to other regions through time.

V. REFERENCES

Abramovsky, L., Rupert Harrison, and Helen Simpson. "University Research and the Location of Business R & D," *The Economic Journal*, 117, 2007, C114–C141.

Acs, Zoltan J., Luc Anselin, and Attila Varga. "Patents and Innovation Counts as Measures of Regional Production of New Knowledge." *Research Policy*. 3, 2002, 1069-1085.

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

Acs, Zoltan J., David B. Audretsch, and Maryann P. Feldman. "Real Effects of Academic Research: Comment." *American Economic Review*, 82, 1992, 363-367.

Anselin, Luc, Attila Varga, and Zoltan J. Acs. "Geographical Spillovers and University Research: A Spatial Econometric Perspective." *Growth and Change*, 31, 2000, 501-515.

Audretsch, David B. and Maryann P. Feldman. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review*, 86, 1996, 630-640.

Audretsch, David B. and Talat Mahmood. "New Firm Survival: New Results Using a Hazard Function." *The Review of Economics and Statistics*, 77, 1995, 97-103.

AUTM (2004), Licensing Survey, available at http://www.autm.org/events/File/FY04%20Licensing%20Survey/04AUTM USLicSrvy-public.pdf.

Baldwin, John R. and Paul K. Gorecki. "Firm Entry and Exit in the Canadian Manufacturing Sector, 1970-1982." *The Canadian Journal of Economics*, 24, 1991, 300-323.

Cameron, Colin A., and Pravin K. Trivedi. "*Microeconometrics: Methods and applications*." Cambridge University Press, NY, 2005.

Chapple, Karen, Ann Markusen, Greg Schrock, Daisaku Yamamoto, and Pingkang Yu. "Gauging Metropolitan High-Tech and "I-Tech" activity," *Economic Development Quarterly*, 18, 2004, 10-29.

Dunne, T., S. D. Klimek, and Mark J. Roberts. "Exit from regional manufacturing markets: The role of entrant experience," *International Journal of Industrial Organization*, 23, 2005, 399.421.

Dunne, Timothy, Mark J. Robert, and Larry Samuelson. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *The Rand Journal of Economics*, 19, 1988, 495-515.

Dunne, Timothy, Mark J. Robert, and Larry Samuelson. "The Growth and Failure of U.S. Manufacturing Plants." *The Quarterly Journal of Economics*. 104, 1989, 671-698.

Dunne, Timothy, Mark J. Robert, and Larry Samuelson. “Firm Entry and Postentry Performance in the US Chemical Industry.” *Journal of Law and Economics*, 32, 1989, 233-271.

Fischer, Manfred M., and Attila Varga. “Spatial Knowledge Spillovers and University Research: Evidence from Austria.” *The Annals of Regional Science*, 37, 2003, 303-322.

Gourieroux, Christian, Alain Montfort, and Alain Trognon. “Pseudo Maximum Likelihood Methods: Theory.” *Econometrica*, 52, 1984a, 681-700.

________“Pseudo Maximum Likelihood Methods: Application to Poisson Models.” *Econometrica*, 52, 1984b, 701-721.

Guimaraes, Paulo. “The fixed effects negative binomial model revisited.” Economics Letters, 99, 2008, 63-66.

Guimaraes, P, and R. Lindrooth. “Controlling for Overdispersion in Grouped Conditional Logit Models: A Computationally Simple Application of Dirichlet Multinomial Regression.” *Econometrics Journal*, 10, 2007, 439-452.

________“Dirichlet-Multinomial Regression.” RePEc Working Paper # 0509001.

Jaffe, Adam B. “Real Effects of Academic Research.” *American Economic Review*, 79, 1989, 957-970.

Jaffe, Adam B., Manuel Trajtenberg,and Rebecca Henderson. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *The Quarterly Journal of Economics*, 108, 1993, 577-598.

Mansfield, Edwin. “Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing.” *Review of Economic and Statistics*, 77, 1995, 55-65

Paytas, Jerry and Dan Berglund. “Technology Industries and Occupations for NAICSIndustry Data.” Pittsburgh: Carnegie Mellon Center for Economic Development and State Science and Technology Institute.2004. http://www.ssti.org/Publications/Onlinepubs/NAICS_Tech1.pdf

Paytas, Jerry, Robert Gradeck, and Lena Andrews. “Universities and the Development of Industry Clusters”, Carnegie-Mellon Center for Economic Development, U.S. Economic Development Administration, 2004.

Porter, Michael. “Clusters and Competition: New Agendas for Companies,Governments, and Institutions” in *On Competition*, Harvard Business School, 1998.

Wallsten, Scott J. “An Empirical Test of Geographic Knowledge Spillovers using Geographic Information Systems and Firm-size-level Data.” *Regional Science and Urban Economics*, 2001, 571-599.

Woodward, Douglas, Octávio Figuerido, and PauloGuimarães. "Beyond the Silicon Valley: Univeristy R&D and high technology location." *Journal of Urban Economics*, 60, 2006, 15-32.

Wooldridge, Jeffrey, "*Econometric Analysis of Cross Section and Panel Data*," MIT Press,Cambridge, 2002.

Zucker, Lynne G., Michael R. Darby, and Marilynn B Brewer. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, 88, 1998, 290-306.

Figure 1: High-tech firm locations in Texas

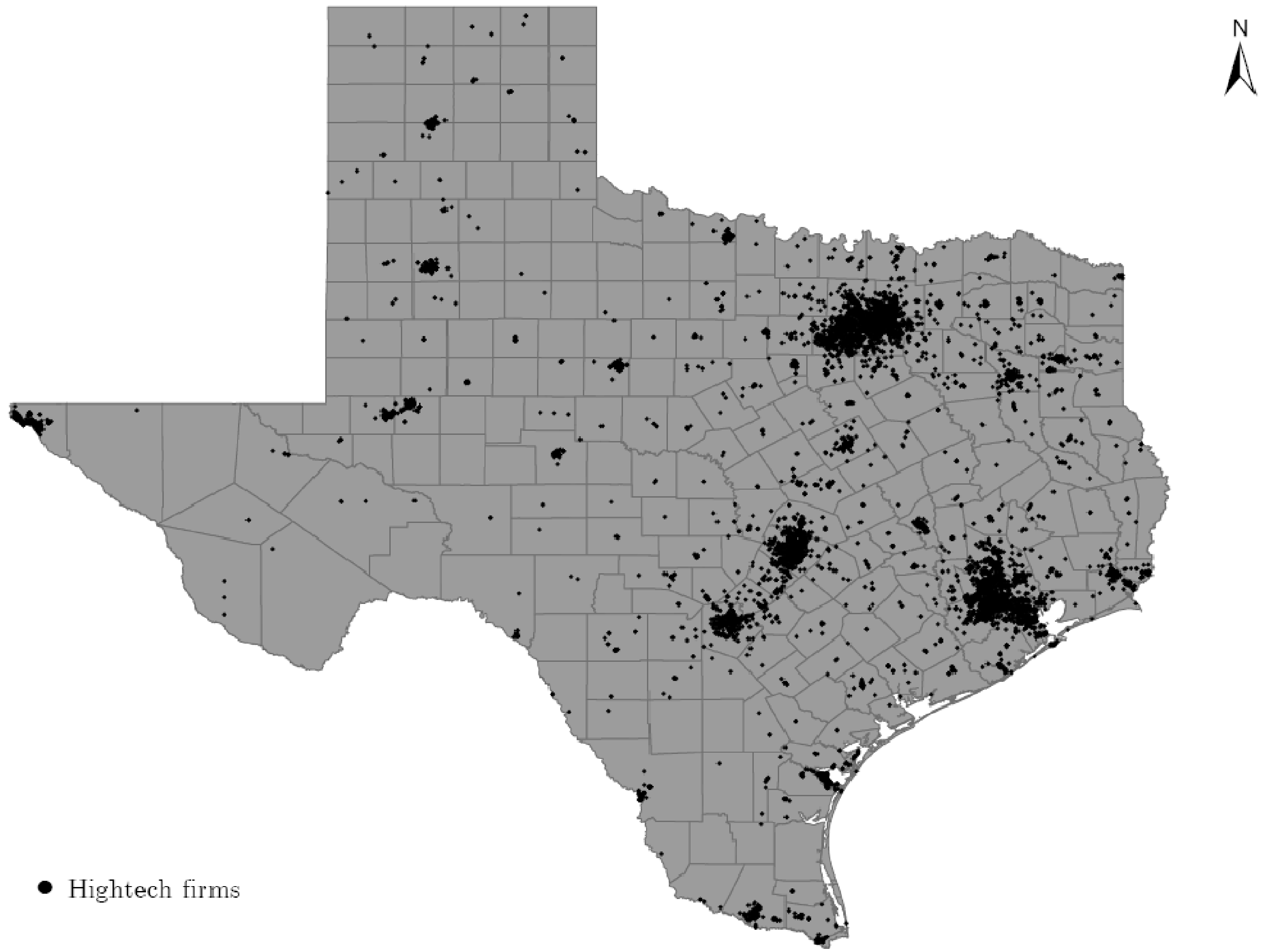


Table 1: Start-ups' and Incumbents' Summary Statistics

Variable	All Counties		MSA Counties Only		Knowledge Center Counties	
	Start-ups	Incumbents	Start-ups	Incumbents	Start-ups	Incumbents
Average number	.065	.683	.205	2.123	1.186	12.347
of firms	(.993)	(11.071)	(1.791)	(20.027)	(4.694)	(53.543)
Average number	.063	.110	.200	.349	1.211	2.126
of exits	(1.393)	(2.613)	(2.522)	(4.735)	(6.796)	(12.833)
Average number	39.904	49.810	41.314	52.019	42.090	59.982
of Employees	(363.975)	(346.560)	(373.666)	(356.705)	(389.482)	(401.247)
Average income	5,185.20	5,493.39	5,251.04	5,604.26	5,502.64	5,866.79
	(5,263.73)	(7,693.12)	(5,349.74)	(7,871.16)	(5,792.92)	(7,101.32)

Standard errors are in parenthesis.

Table 2: Funding by County

County	Fiscal Year								
	1995	1996	1997	1998	1999	2000	2001	2002	2003
Bexar	132,240	143,603	123,456	131,522	131,591	166,159	180,733	168,448	202,852
Brazos	169,888	147,948	160,234	131,986	180,576	187,164	239,402	234,820	258,288
Dallas	193,532	378,111	406,623	408,488	328,480	258,363	300,453	633,820	516,580
Denton	8,175	6,509	6,843	7,848	18,030	19,569	23,471	22,420	29,283
El Paso	21,672	35,399	32,564	16,589	35,209	20,900	25,856	30,721	27,733
Galveston	85,920	93,985	102,311	101,200	162,139	179,442	197,201	226,632	356,799
Harris	602,505	647,880	607,783	736,785	752,151	915,044	1,107,989	1,317,971	1,344,938
Lubbock	18,067	20,765	22,934	27,828	27,975	38,817	42,006	38,814	52,449
Tarrant	82,676	37,346	34,569	26,283	15,939	11,445	14,033	23,051	28,896
Travis	280,091	351,267	271,068	258,116	273,261	316,759	364,421	305,650	505,583

In '000 of 1999:3 dollars

Table 3: Average Spillover and Distance to the Closest Knowledge Center

County	Average distance	Spillover = Funding (in '000 of dollars)/ Distance			
		Lagged by one year	Lagged by two years	Lagged by three years	Lagged by four years
Bexar	224.243	1,063.84	977.77	950.83	898.26
	(106.016)	(977.35)	(895.59)	(869.26)	(812.51)
Brazos	214.837	1,398.07	1,286.82	1,188.25	1,109.39
	(97.025)	(1,592.01)	(1,461.50)	(1,347.21)	(1,229.97)
Dallas	217.155	2,775.09	2,650.35	2,360.26	2,238.96
	(105.507)	(2,818.64)	(2,679.20)	(2,225.40)	(2,181.58)
Denton	221.808	129.84	105.68	88.56	72.09
	(101.041)	(135.17)	(115.48)	(103.11)	(83.26)
El Paso	481.598	64.14	66.11	68.02	66.31
	(149.829)	(49.09)	(50.91)	(52.92)	(52.26)
Galveston	297.771	999.30	791.43	683.08	592.19
	(149.312)	(1,038.60)	(776.63)	(674.04)	(584.49)
Harris	289.235	6,200.73	5,460.49	4,787.60	4,280.00
	(145.521)	(8,710.00)	(7,746.46)	(6,670.33)	(5,864.01)
Lubbock	275.719	223.20	194.29	176.62	153.17
	(135.653)	(277.28)	(240.67)	(222.45)	(192.78)
Tarrant	207.820	138.41	144.98	161.51	240.92
	(96.8559)	(138.89)	(150.58)	(173.74)	(308.13)
Travis	204.445	2,361.19	2,087.58	2,140.80	2,042.41
	(101.141)	(2,159.98)	(1,817.86)	(1,871.83)	(1,773.69)

Standard errors are in parenthesis. Data based on values in Table 2 and distance in miles.

Table 4: Summary Statistics for the Regression Variables

Variable	Mean (Standard Deviation)
Number of start-ups by industry	.065
	(.993)
Log of Funding relative to distance to county	
Spillover lagged by one year	8.442
	(.804)
Spillover lagged by two years	8.328
	(.809)
Spillover lagged by three years	8.237
	(.799)
Spillover lagged by four years	8.147
	(.796)
Base county's	
Knowledge center county	.039
	(.194)
MSA county	.303
	(.460)
Log of lagged employment of start-ups by industry	.041
	(.342)
Log of lagged employment of incumbents by industry	.189
	(.753)
Lagged number of exits by start-ups by industry	.047
	(1.262)
Lagged number of exits incumbents by industry	.094
	(2.531)
Income (in '000)	2.425
	(.986)
Amenity	3.151
	(2.428)
Land price (in '00)	2.680
	(2.126)
Percentage of population between ages 20 and 44	40.007
	(5.214)
Neighboring counties'	
Log of lagged employment of start-ups by industry	.080
	(.407)
Log of lagged employment of incumbents by industry	.361
	(.875)
Lagged number of exits by start-ups by industry	.047
	(.777)
Lagged number of exits incumbents by industry	.094
	(1.550)
Amenity	3.821
	(2.166)
Percentage of population between ages 20 and 44	41.979
	(4.512)
Time and business variables	
FY 2000-01, FY 2001-02, FY 2002-03, FY 2003-04, and FY 2004-05.	.167
	(.373)
Unemployment rate (county level)	5.551
	(1.926)
Prime lending rate	5.547
	(.826)

Standard deviations are in parentheses

Table 5a: Start-up Patterns for Technology Firms (Fixed Effects by Six-digit NAICS Codes)

Variable	Number of new start-ups for a county per year			
	(1)	(2)	(3)	(4)
Log of Funding relative to distance to county				
Spillover lagged by one year	.225***	.		
	(.037)			
Spillover lagged by two years		.232***		
		(.038)		
Spillover lagged by three years			.243***	
			(.040)	
Spillover lagged by four years				.217***
				(.038)
Base county				
Knowledge center county	1.379***	1.378***	1.377***	1.382***
	(.061)	(.061)	(.061)	(.061)
MSA county	1.547***	1.538***	1.531***	1.540***
	(.092)	(.090)	(.088)	(.089)
Log of lagged employment of start-ups by industry	.251***	.262***	.262***	.255***
	(.054)	(.057)	(.060)	(.057)
Log of lagged employment of incumbents by industry	.355***	.360***	.361***	.358***
	(.079)	(.081)	(.083)	(.083)
Lagged number of exits by start-ups by industry	-.012	-.015*	-.019**	-.017*
	(.008)	(.008)	(.008)	(.009)
Lagged number of exits incumbents by industry	.003	.004	.006	.005
	(.004)	(.004)	(.004)	(.004)
Income (in '000)	.348***	.346***	.343***	.348***
	(.032)	(.032)	(.031)	(.031)
Amenity	.181***	.180***	.180***	.179***
	(.018)	(.019)	(.019)	(.019)
Land price (in '00)	.084***	.085***	.087***	.087***
	(.009)	(.009)	(.009)	(.009)
Percentage of population between ages 20 and 44	.046***	.047***	.048***	.048***
	(.007)	(.008)	(.008)	(.008)
Unemployment rate	-.142***	-.142***	-.143***	-.149***
	(.046)	(.046)	(.046)	(.047)
Neighboring counties'				
Log of lagged employment of start-ups by industry	-.075	-.089	-.091	-.084
	(.057)	(.058)	(.063)	(.060)
Log of lagged employment of incumbents by industry	-.071	-.080	-.082	-.079
	(.090)	(.093)	(.095)	(.096)
Lagged number of exits by start-ups by industry	.027*	.031**	.039**	.035**
	(.014)	(.014)	(.015)	(.015)
Lagged number of exits incumbents by industry	.000	-.001	-.004	-.004
	(.008)	(.007)	(.007)	(.007)
Amenity	-.126***	-.122***	-.121**	-.122**
	(.028)	(.028)	(.028)	(.028)
Population percentage between ages 20 and 44	-.017*	-.019*	-.021*	-.021**
	(.009)	(.010)	(.010)	(.010)
Number of obs.	76200	76200	76200	76200
Wald χ^2	17273.44	17301.14	17303.91	17281.73

***Denotes statistical significance at the 1% level, ** denotes statistical significance at the 5% level and * denotes statistical significance at the 10% level. We have included five year dummies.

Table 5b: Start-up Patterns for Technology Firms in TX Using Dirichelet-Multinomial Model (Group Effects by Six-digit NAICS Codes)

Variable	Number of new start-ups for a county per year			
	(1)	(2)	(3)	(4)
Log of Funding relative to distance to county				
Spillover lagged by one year	.242*** (.034)			
Spillover lagged by two years		.252*** (.033)		
Spillover lagged by three years			.262*** (.033)	
Spillover lagged by four years				.235*** (.033)
Number of obs.	76200	76200	76200	76200
Wald χ^2	9501.75	9547.03	9571.36	9552.30

***Denotes statistical significance at the 1% level, ** denotes statistical significance at the 5% level and * denotes statistical significance at the 10% level. We have included five year dummies.

Table 6: Kaplan-Meier Survival Function Estimates for Start-ups and Incumbents

Time	Start-ups	Incumbents
3 months	.942	.999
	(.003)	(.000)
6 months	.872	.999
	(.004)	(.000)
12 months	.758	.993
	(.005)	(.001)
24 months	.495	.964
	(.006)	(.002)
36 months	.360	.917
	(.006)	(.003)

Standard errors are in parenthesis.

Table 7: Cox Proportional Hazard Model Estimates for New Start-ups and Incumbents

Variable	(1)	(2)	(3)	(4)
Start-up firm	2.396***	2.367***	2.838***	2.829***
	(.126)	(.126)	(.194)	(.195)
Spillover dummy (*25%-50%*)	-.085	-.085	-.040	-.041
	(.150)	(.150)	(.162)	(.162)
Spillover dummy (*50%-70%*)	.016	.016	.035	.037
	(.088)	(.088)	(.081)	(.081)
Spillover dummy (*75%-100%*)	.035	.041	.094	.102
	(.134)	(.135)	(.122)	(.122)
Spillover dummy (*25%-50%*) ×	.008	.011	-.035	-.033
Start-up firm	(.156)	(.157)	(.169)	(.169)
Spillover dummy (*50%-75%*) ×	.008	.017	-.019	-.012
Start-up firm	(.095)	(.095)	(.088)	(.088)
Spillover dummy (*75%-100%*) ×	-.081	-.079	-.142	-.144
Start-up firm	(.146)	(.146)	(.133)	(.133)
Distance to nearest knowledge	-.004**	-.004**		
center	(.002)	(.002)		
Distance to nearest knowledge	.003	.003		
center × *Start-up firm*	(.002)	(.002)		
≤ 50 miles			.431**	.444**
			(.194)	(.194)
>50 - ≤ 75 miles			.354	.366
			(.236)	(.236)
>75 - ≤ 100 miles			-.196	-.187
			(.327)	(.328)
≤ 50 miles ×			-.311	-.323
Start-up firm			(.204)	(.205)
>50 - ≤ 75 miles ×			-.287	-.290
Start-up firm			(.250)	(.251)
>75 - ≤ 100 miles ×			.173	.167
Start-up firm			(.343)	(.343)
Employment ratio	-.415***	-.421***	-.374***	-.379***
	(.068)	(.069)	(.070)	(.070)
Log of average wage	-.201***	-.204***	-.203***	-.206***
	(.021)	(.021)	(.021)	(.021)
Unemployment rate	.506***		.505***	
	(.023)		(.023)	
Prime rate		-.252***		-.252***
		(.019)		(.019)
# of Observations	17152	17152	17152	17152
Wald χ^2	5224.316	4680.739	5220.528	4677.108

Robust standard errors are in parentheses. ***Denotes statistical significance at the 1% level, ** denotes statistical significance at the 5% level and * denotes statistical significance at the 10% level.

Figure: 2 Hazard Rates for all Start-ups and Incumbents

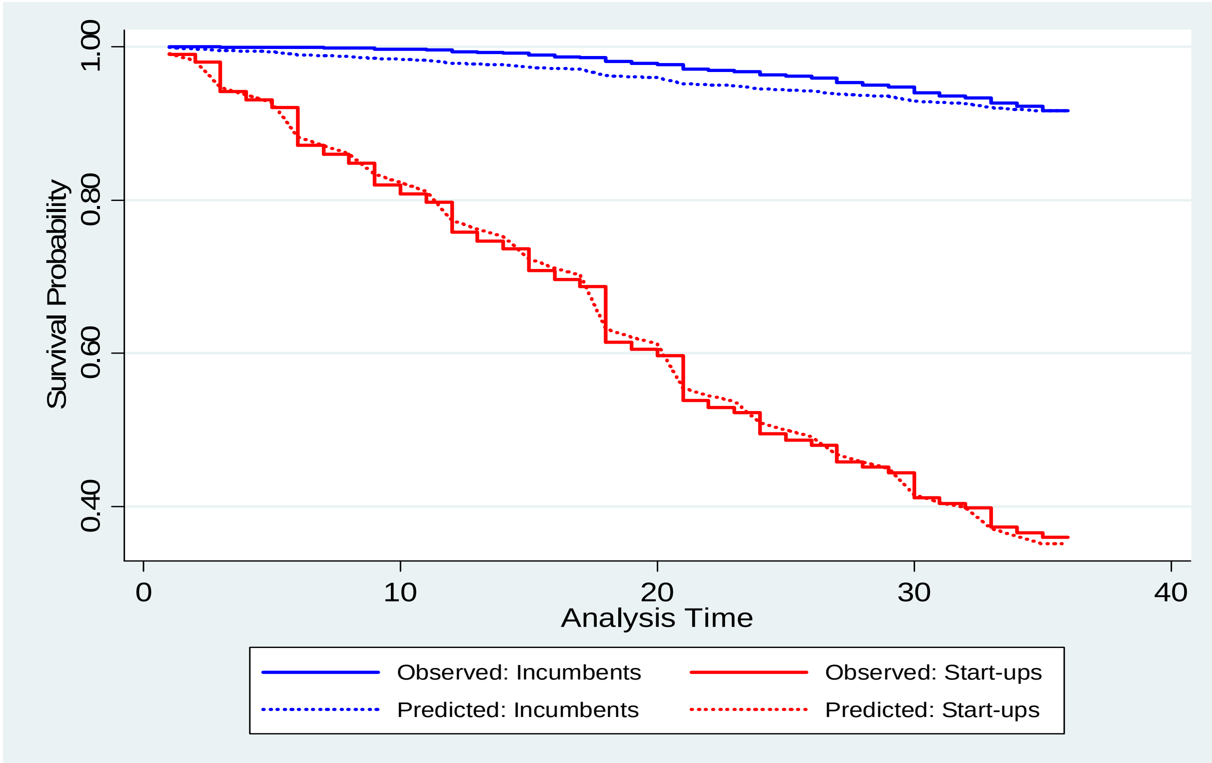


Table 8: Distribution of Start-ups and Incumbents by Distance to Knowledge centers

Distance	Start-ups	Incumbents
≤ 50 miles	6479	7866
>50 - ≤ 75 miles	513	744
>75 - ≤ 100 miles	278	297
> 100 miles	443	532

Figure 3: Hazard Rates for New Start-ups and Incumbents by Spillover

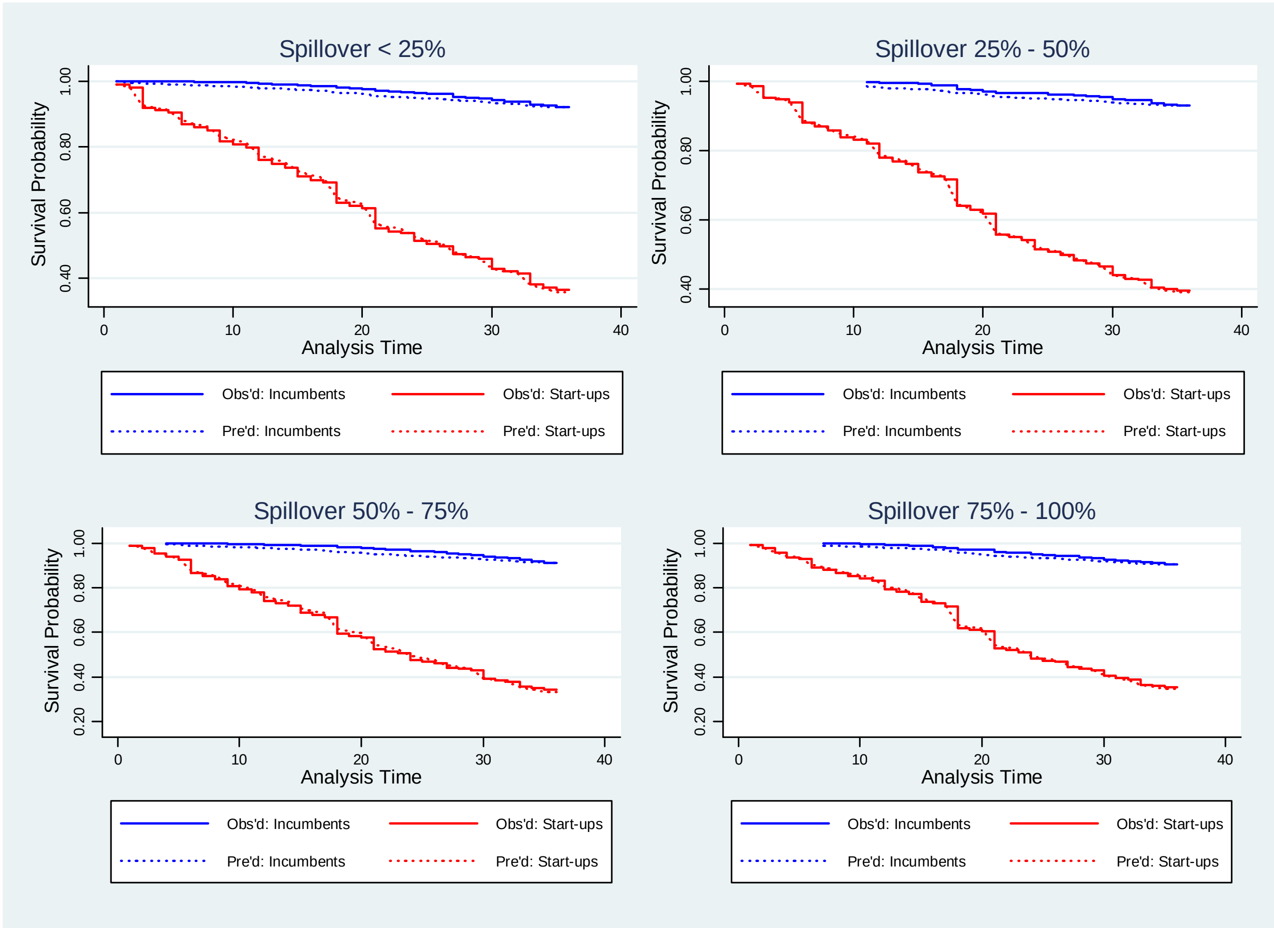


Figure 4: Hazard Rates for New Start-ups and Incumbents by Distance

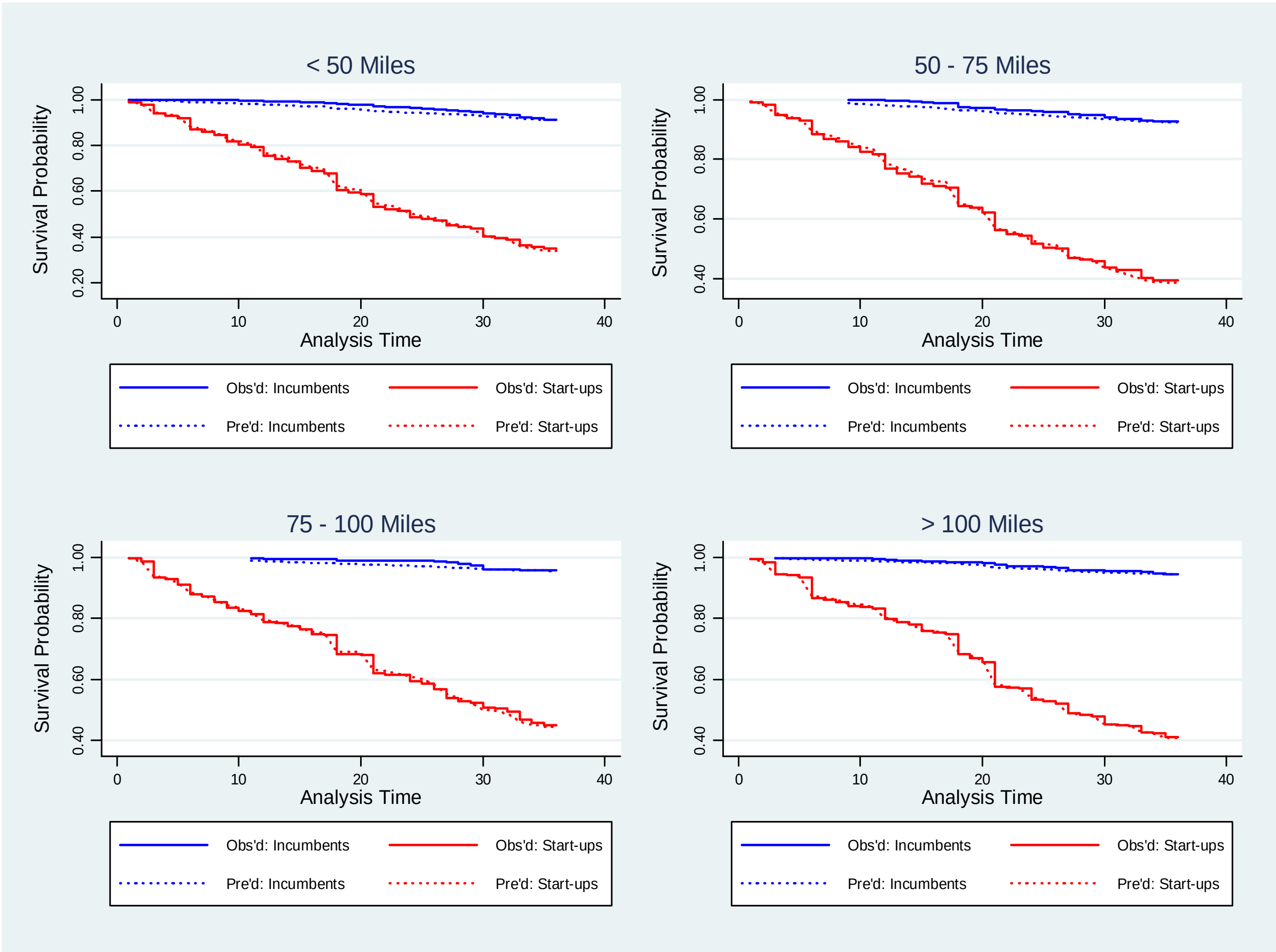
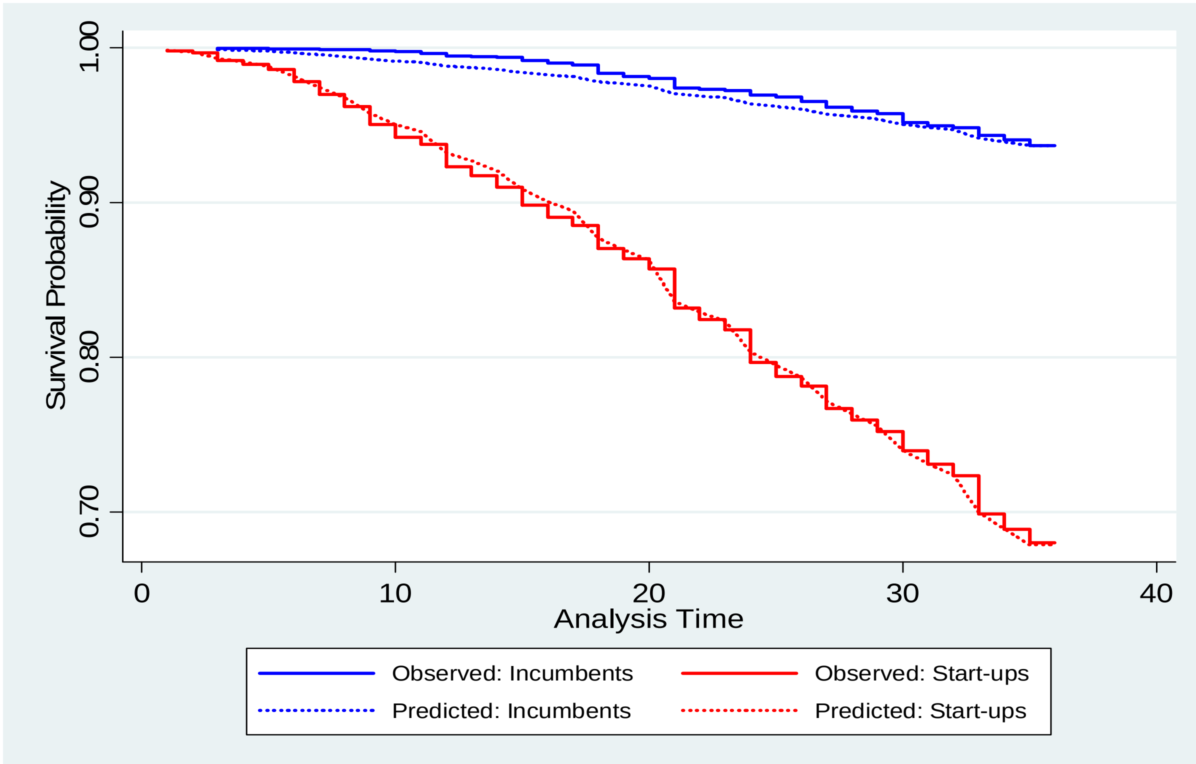


Figure 5: Hazard Rates for all Start-ups and Incumbents with Alternative Definition



VI. APPENDIX

Table A1: High-Tech Industry Classifications

NAICS	Description	NAICS	Description
	Defense Electronics		**Measuring & Control Instruments**
334511	Search, Detection, Navigation, Guidance, Aeronautical, and	334512	Automatic Environmental Controls
	Nautical Systems and Instruments	334513	Industrial Process Control Instruments
	Computer & Peripheral Equipment	334514	Totalizing Fluid Meter & Counting Devices
334111	Electronic Computers	334515	Electricity Measuring & Testing Equipment
334112	Computer Storage Devices	334516	Analytical Laboratory Instruments
334113	Computer Terminals[a]	334519	Other Measuring & Controlling Instruments Electromedical
334119	Other Computer Peripheral Equipment		Equipment
	Communications Equipment	334510	Electromedical & Electrotherapeutic Apparatus
334210	Telephone Apparatus	334517	Irradiation Apparatus[a]
334220	Radio & TV Broadcasting & Wireless Communications		**Communications Services**
	Equipment	517110	Wired Telecommunications Carriers
334290	Other Communications Equipment	517211	Paging Services
335921	Fiber Optic Cables	517212	Cellular & Other Wireless Telecommunications
	Consumer Electronics	517310	Telecommunications Resellers
334310	Audio & Video Equipment	517410	Satellite Telecommunications
	Electronic Components	517510	Cable & Other Program Distribution
334411	Electron Tubes	517910	Other Telecommunications
334412	Bare Printed Circuit Boards		**Computer Systems Design & Related Services**
334414	Electronic Capacitors	541511	Custom Computer Programming
334415	Electronic Resistors	541512	Computer Systems Design
334416	Electronic Coils, Transformers, & other Inductors	541513	Computer Facilities Management
334417	Electronic Connectors	541519	Other Computer Related Services
334418	Printed Circuit Assembly		**Internet Services**
334419	Other Electronic Components	518111	Internet Service Providers
	Semiconductors	518112	Web Search Portals
334413	Semiconductor & Related Devices[a]	518210	Data Processing, Hosting, & Related Services
333295	Semiconductor Machinery		**R&D & Testing Labs**
	Photonics	541710	Research & Development in the Physical, Engineering, & Life Sciences
333314	Optical Instrument & Lens	541380	Testing Laboratories
333315	Photographic & Photocopying Equipment		**Biotechnology Companies**
	Software Publishers	325411	Medicinal Chemicals and Botanical Products
511210	Software Publishers	325412	Pharmaceutical Preparations
	Engineering Services	325413	In Vitro and In Vivo Diagnostic Substances
541330	Engineering Services	325414	Biological Products, Except Diagnostic Substances
	Computer Training	541711	Commercial Physical and Biological Research[a]
611420	Computer Training		

a: These NAICs codes are not used in this study due to lack of variation.

Table A2: Academic Institutions by County

County	University	County	University
Bexar	Trinity University	*Harris*	Baylor College of Medicine
	U of TX Health Science Center		Houston Community College
	U of TX San Antonio		Prairie View A&M University
	U of the Incarnate Word		Rice University
Brazos	Alvin Community College		San Jacinto College District system
	Brazosport College		Texas Southern University
	TX A&M U. System Health Science Center		U of TX Health Science Center
	Texas A&M U System Office		U of TX MD Anderson Cancer Center
	Texas A&M University		U of Houston
Dallas	Paul Quinn College		
			U of Houston Clear Lake
	Richland College		U of Houston System Administration
	Southern Methodist University		U of St. Thomas
	U of TX South-West Medical Center	*Lubbock*	Lubbock Christian University
	U of Dallas		Texas Tech University
	U of TX Dallas	*Tarrant*	Texas Wesleyan University
Denton	U of North Texas		U of TX Arlington
El Paso	U of TX El Paso		U of North Texas Health Science Center
Galveston	College of the Mainland	*Travis*	Austin Community College
	U of TX Medical Branch		St Edwards University
			U of TX Austin
			U of TX System Office

Endnotes

[i] For a description of the cluster concept as adopted by economic development practitioners, see Micahel Porter (1998.)
[ii] Also see Audretsch and Feldman (1996) where they examined the link between knowledge spillovers in an industry and geography of innovation and production.
[iii] For a survey of university technology transfer efforts, see Paytas et al, (2004.)
[iv] See AUTM 2004 Licensing Survey.
[vii] This is based on the BLS Occupational Employment Statistics at the 3 digit level of aggregation and a subset of occupations designated as science and engineering intensive by Chapple et al (2004.)
[viii] Even at the NAICS-2 levels containing the high-tech industries, some counties have zero employment. In these cases, bearing in mind that we are trying to capture wage and income levels in higher skill activities, we compute an average wage using NAICS 31-33 (Manufacturing), 51(Information), 52(Finance and Insurance), 54(Professional, Scientific, and Technical Services), 55(Management of Companies and Enterprises), 61(Educational Services), & 62 (Health Care and Social Assistance). A relatively large share of employment in these industries requires a bachelor's degree.
[ix] Four industries at 6-digit NAICS were omitted since there was no variation in start-ups over the period of this analysis, (see Table A1), so there are only 50 industries being tracked. The total number of observations is therefore 50 industries multiplied by 254 counties multiplied by six years (1999:3 – 2005:2) or a total of 76,200 observations.
[x] All ten knowledge center counties are MSA counties, but not all MSA counties are knowledge center counties.
[xi] We set this requirement in nominal dollars although R&D expenditures are deflated to 1993:3 in the regression analysis. As is evident from Table 2, deflating R&D expenditures to 1999:3 will not change the set of research centers.
[xii] The TWC unemployment rate is the average rate for the calendar year. We average consecutive years beginning with year 1999-2000 since that best overlaps our definition of a year as running from third quarter through second quarter of the following calendar year.
[xiii] Also note that Guimaraes (2008) show that that the conditional maximum likelihood estimator of the negative binomial with fixed effects does not necessarily remove the individual fixed effects in count panel data. This will happen only if the number of groups is at least 1000 with more than 20periods per group.
[xiv] The negative binomial model, by contrast, is only consistent if the conditional distribution of the dependent variable is in fact a negative binomial. Thus, the potential benefit from preferring the negative binomial over Poisson, i.e., increased efficiency if the data are over-dispersed, must be balanced against the more restrictive conditions that need to be met to ensure consistency.
[xv] For a more detailed discussion of this reasoning, see Wooldridge (2002) and Cameron and Trivedi (2005).
[xvii] This actual variable is of course employment plus one.
[xviii] A list of MSA counties in Texas can be found at http://txsdc.utsa.edu/tpepp/msa04_list.php
[xix] We also use relative spillover as a variable. Relative spillover for firm *i* is calculated as the average spillover for given firm *i* divided by the average spillover for all firms.
[xx] We used state-level seasonally unadjusted monthly unemployment rates reported by the BLS.