Industrial Agglomeration and Spatial Persistence: Entry, Growth, and Exit of Software Publishers

Deltas, George and De Silva, Dakshina G. and McComb, Robert P.

University of Illinois, Lancaster University, Texas Tech University

22 September 2014

Industrial Agglomeration and Spatial Persistence: Entry, Growth, and Exit of Software Publishers*

George Deltas[†] Dakshina G. De Silva[‡] Robert P. McComb[§]

October 29, 2014

Abstract

We use geocoded administrative data from Texas on all business establishments to estimate the effects of localization economies on the spatial persistence of industrial employment for the software industry. We decompose this persistence into components arising from entry rates, firm growth, and exit rates. Unlike previous research that has used geographies based on county and MSA divisions, this analysis takes place at a very high level of spatial resolution in which the industrial composition is identified within areas as small as one mile in radius. The choice of the software industry allows us to isolate the effects arising from human capital spillovers and the effects arising from the labor pool channel from other sources of agglomeration economies. Moreover, the decomposition of the employment persistence in entry, growth and exit, and the high level of spatial resolution allow us to distinguish between these two effects and has a number of other advantages. The results suggest that a location, defined as a 1-mile radium circle, with an initial concentration of software industry employment, retains a disproportionate number of software industry employees 6 years later. Software industry employment in the surrounding area has a small and often insignificant effect, i.e., any agglomeration effects dissipate rapidly over space. The results are not driven by higher growth rates of software establishments in high concentration locations or by differences in the survival probabilities. Rather, they are fully accounted for by two factors: (i) the retention of jobs lost by an establishment in a location by other establishments in that same location and (ii) an increased propensity of software establishments to enter in or near locations with prior software establishment presence. The entry effect diminishes sharply beyond one mile. These findings are mostly consistent with labor channel effects, including the possibility of spin-offs locating near existing firms, but disembodied human capital spillovers might also be present to some extent.

JEL Classification: R12.
Keywords: Agglomeration economies, labor pools, knowledge spillovers, firm growth.

*This paper replaces earlier research that was circulated under a similar title. For that earlier research, we want to thank Tim Dunne and seminar participants in the Dusseldorf Center for Industrial Economics (DICE), the University of Illinois, the University of Lancaster, University of Macedonia, the EARIE conference, the International Industrial Organization conference, and the Western Economic Association conference for their helpful comments. We also would like to thank Anita Schiller, and Mervin Ekanayake for their skillful research assistance, the Texas Workforce Commission for providing the data, and a number of sources with experience in the software industry.

[†]Department of Economics, University of Illinois, 1407 W. Gregory, Urbana, IL 61801, United States. (e-mail: deltas@illinois.edu)

[‡]Department of Economics, Lancaster University Management School, Lancaster University, Lancaster, LA1 4YX, United Kingdom. (e-mail: d.desilva@lancaster.ac.uk).

[§]Department of Economics, PO Box 41014, Texas Tech University, Lubbock, TX 79409, United States. (e-mail: robert.mccomb@ttu.edu)

1 Introduction

The geographic distribution of high technology firms in the United States differs significantly from the distribution of broader economic activity. Silicon Valley in California, the Route 128 Corridor in Boston, the Research Triangle in North Carolina, and the greater Austin, Texas, region are well known examples of high technology industry concentrations. This phenomenon is, of course, not exclusive to these relatively new industries. Marshall postulated as far back as 1920 that the presence of specialized inputs and upstream service providers, labor pooling, and knowledge spillovers are localized and help to explain why certain industries that are not otherwise tied to geographically specific inputs are spatially concentrated. A voluminous empirical literature has since confirmed the ubiquity of spatial aggregation and investigated the effect of agglomeration spillovers on technical change, firm growth, productivity, and other outcomes. Moreover, spatial aggregation tends to be persistent over time because agglomeration generates self-sustaining dynamics through these positive spillovers.

In this paper, we aim to isolate the importance of agglomeration effects arising from a localized labor pool and from disembodied knowledge spillovers in the spatial persistence of industrial employment. We also aim to identify the mechanics of this persistence by also investigating the extent to which agglomeration impacts establishment entry, employment growth, and exit. To isolate the effects of labor pooling and knowledge spillovers from the many other sources of agglomeration spillovers, we use novel panel data from the software publishing industry that is characterized by a very fine level of spatial granularity.

The software publishing industry lends itself well to the investigation of labor market and knowledge spillovers. The industry's output is primarily intangible intellectual property that sells in the state-wide (if not national or global) market. Thus, locational factors such as access to natural resources, local demand, proximate input suppliers, and transportation costs are not particularly relevant. It is a relatively new industry, so historical factors should not have a significant impact on firm and industry location. Moreover, capital inputs are not typically fixed (beyond the period of the current building lease) or tend to depreciate very rapidly (e.g., computer hardware and software), so past investment decisions in physical plant do not constrain firm location over the medium term. On the other hand, given the highly dynamic and competitive nature of the industry, the rapid evolution of software development, and its heavy reliance on specialized human capital, software publishing seems well suited to benefit from labor pooling and knowledge spillovers. Computer programmers, the most important labor input in this industry, tend to be young and mobile.

Therefore, if labor market driven localization economies are generally present, their effects should be easier to observe in such an industry. The absence of other plausible channels for such spillovers, and our ability to control for localized economic activity given our detailed spatial data on other business establishments (which we describe next) permit us to plausibly isolate these labor channel localization economies from other spillover explanations.

Our identification strategy relies on the quality of our primary administrative data which contain very detailed geographic information and cover the location of every establishment in the State of Texas from the first quarter of 2000 to the end of 2006 on a quarterly frequency. The very fine scale of geographical information permits the identification of spillovers over very small distances and can in principle help distinguish between spillovers through random interaction in the proximity of employment (which may be very localized) with those arising through the labor pool (which might be somewhat less localized). The comprehensiveness of the database permits us to control for overall economic activity at any level of geographical aggregation, removing a possible source of spurious correlation. The panel nature of the data, combined with its geographic detail and comprehensiveness, permits a credible investigation of entry decisions of software publishers. We are thus able to investigate firm dynamics starting from the entry propensities at each location, to establishment growth, and (ultimately) to establishment exit, i.e., we estimate the agglomeration effects at multiple levels of spatial resolution to all parts of a firm's lifecycle. This also provides additional identification power on the nature of these agglomeration effects since we are able to distinguish effects that increase firm productivity (manifested primarily in firm employment growth and increased survival probabilities) from effects that increase industry productivity (manifested primarily in increased entry in a locality and retention of jobs that a firm sheds by other firms in the same locality). This is of course a simplification, since industry and firm productivity are interrelated, and we will use the combination of observed outcomes to infer the relative importance of human capital and labor market spillover effects.

Our starting point is the observation that clusters of software establishments tend to be highly persistent, despite the fact that there is substantial entry and exit in this industry. Moreover, the formation of new software establishment clusters appears rare, even though the industry is expanding and even though there is much expansion in the geographic footprint of economic activity in Texas of industries that use similar infrastructure as software firms. This suggests that an area with an initial presence of software firms tends to maintain that presence over time. In our first major finding, we measure this persistence at the resolution of 1-mile radius. Comparing locations with a positive employment in the software industry, we find that the elasticity of employment at the end of 2006 with respect to employment at the start of 2000 is approximately $0.75 - 0.80$ for reasonable specifications (there is some variation outside this range for different models). These are very high numbers given that the majority of the software jobs (i.e., employment positions rather than employees) that were present at the start of 2000 were lost to other establishments, and given that many new jobs were created by entrants. We emphasize that these figures are obtained after controlling for overall establishment growth at the locality level and other effects.

When estimating this employment elasticity at the firm level, considering for reasons of comparability only firms that were active at the start of the sample, we find that it is lower by about 0.25 for the most directly comparable specifications. We also find that end-of-period employment is lower for an establishment located in areas with prior presence of other software establishments, despite the absence of competition on the product side. Even though this effect is rather weak, it is suggestive that the presence of other software establishments reduces, rather than enhances the success and growth prospects of other incumbent establishments. The difference between the firm level and location level elasticity, and the fact that firm level employment is decreasing in the industry activity in the same location, suggests that many of the jobs that are lost by a software firm in a location are captured by other software firms that are located no more than one mile away. Even this, however, is not sufficient to fully explain the high elasticity at the location level, leaving the location choices of new entrants as the source of the additional jobs.

We then turn to the investigation of entry rates at different locations, and the survival of these entrants as a function of location characteristics (including prior presence of other software establishments). After controlling for localized presence in a control group of other businesses, we find that software firms tend to enter in locations that are within one mile of other software firms. The attractiveness of a location diminishes very rapidly with distance, and almost disappears after 5 miles. The presence of neighboring firms has no effect on exit at any distance threshold. In other words, the high spatial persistence of employment in the software industry is driven by two components: (i) the propensity of software establishments to locate in areas with pre-existing software establishments, and (ii) the ability of co-located firms to act as a "sponge" for jobs lost by their neighbors. The spatial persistence is not driven by differential growth rates of establishments located in areas with substantial software firm presence or by differential survival probabilities of those establishments.

This evidence, in conjunction with weak spatial specialization of software firms (we show that collocated software establishments do not appear to be in the same software business segments), suggests that there is a substantial spatial link through the labor market and a possibly smaller spatial link through human capital (knowledge) spillovers. Close proximity facilitates the transmission of information about the quality of employees of other firms and their match value for positions available with a prospective employer. Physical proximity likely also reduces the switching costs, as employees would not need to make substantive adjustments in living and commuting arrangements. These factors reduce the set-up costs of establishments, and may also somewhat reduce subsequent recruiting costs. As a result, potential entrants tend to choose a location near other existing firms and departing employees of a firm tend to locate a spin-off in the same location. Knowledge spillovers could be an alternative contributory factor in the increased propensity of firms to locate close to other software firms, and explain why locations with an existing agglomeration of software publishers can build on that initial advantage. However, knowledge spillovers on their own could not account for the full range of outcomes we document. In particular, they could not explain why establishment growth and exit rates are unaffected by the close proximity of other software establishments. Indeed, as mentioned above, an increase in the software firm presence in very close proximity tends to marginally decrease employment growth at the firm level, even though it seems to increase employment at the industry level (beyond what would be expected by mere employment inertia). The presence of limited knowledge spillovers is not inconsistent with the presence of larger labor market effects. Indeed, as we explain in detail later in the paper, it might help to explain why the increased entry in high agglomeration locations of some (presumably) marginal firms does not depress the average survival rate.

As noted above, economists have long recognized that industries tend to concentrate spatially when agglomeration economies are present. Agglomeration economies are cast in terms of urbanization economies as described by Henderson (1986) and Krugman (1991), and in terms of own-industry localization economies in the Marshallian vein. Urbanization economies are present when firms in different industries cluster both to take advantage of high levels of demand while also contributing to increased demand. Relatedly, Glaeser *et al.* (1992) refer to the Jacobs-type of urbanization externality that arises when industrial diversity enhances opportunities for inter-industry knowledge spillovers.[1] Agglomeration economies (including knowledge spillovers) have been examined from the perspectives of firm location choices (Rosenthal and Strange (2003), Woodward et al. (2006)), firm exits (Staber, 2001), industry growth (Glaeser et al. (1992), Henderson et al. (1995), and Combes (2000)), and labor productivity (Ciccone and Hall (1996)). Glaeser et

[1]Their reference on this is Jane Jacobs, The Economy of Cities, New York: Vintage 1969.

al. (1992) and Henderson et al. (1995) distinguish between the concepts of static versus dynamic externalities. While the notion of a static agglomeration externality can explain the location of industries, industrial growth suggests a dynamic character to the externality. As articulated by Henderson et al. (1995), dynamic externalities arise as information and experience accumulate through time within the locality. As the concentration and urban density increases through time, there is a deepening in the specialized labor pool and input suppliers while labor mobility and matching between firms is enhanced. Glaeser et al. (1992) introduce the term Marshall-Arrow-Romer externality to describe the dynamic localization economy. Combes (2000) notes that a greater number of similar firms in a locality increases the likelihood of knowledge spillovers since there is greater likelihood of closer matches between firms. This would be the case if knowledge is more readily communicated among people in close proximity. Indeed, since knowledge –as distinct from information– gleaned from experience tends to be embodied in the individual, worker mobility and direct personal and professional interaction are probably the primary channels by which unpriced knowledge spillovers are conveyed. Close geographic proximity should facilitate those channels. Consistent with both the labor pooling and knowledge spillover hypotheses, Audretsch and Feldman (1996) find that industries in which R&D and skilled labor are more important are indeed more concentrated spatially. Moreover, Freedman (2008) finds evidence that spatial clustering facilitates localized worker mobility in the software publishing industry. Rosenthal and Strange (2003) report a quote from Saxenian (1994) in which a high-tech worker from Silicon Valley states, "The joke is that you can change jobs and not change parking lots." For such localized job changes, search and transactions costs are probably negligible.

This latter quote has implications for the appropriate geographical area over which the distinction between the Marshall-Arrow-Romer and Jacobs externalities can be observed. Rosenthal and Strange (2003) and Wallsten (2001) note that localization externalities and knowledge spillovers attenuate rapidly within one mile. Centripetal forces due to the presence of these externalities may drive spatial concentrations in sub-metropolitan pockets within relatively diverse cities. A large and diverse city or metropolitan area may be too large of a geographic division to find evidence of spatial specialization that has occurred as a consequence of Marshall-Arrow-Romer externalities. If so, industry growth from this type of externality may be erroneously attributed to the presence of Jacobs externalities.

This literature shapes our own conceptual framework and the sources of variation in the data on which we rely to distinguish between the different sources of agglomeration spillovers. Our identification and estimation strategy is given in the next section, and is followed by a section that details our data. Section

4 presents the results, while the implications of our findings are summarized in the conclusion.

2 Conceptual Framework and Identification

There are two major possible spatial effects that are likely to be operative for the software publishing industry. The first is the knowledge spillover effect. Workers who work in different firms can learn from each other because of random interactions, either outside of work or as part of business collaboration. These interactions can be of a very fine geographical scale. For example, workers who happen to meet each other at other venues (social or professional gatherings) can far more easily meet (e.g, for lunch, coffee, or even car-pooling) if their places of employment are located very close to each other. Being in the same business cluster, where one might even walk to a meeting, reduces the barriers for such meetings. Interactions are far harder to arrange if the location of employment is even five miles away. The net knowledge spillover is expected to be positive, but there may be "winners" and "losers" for any specific transaction. For example, when one firm learns an idea from another, it may well be to the detriment of the source firm if there is competition between the firms.[2]

The second spatial effect operates through the labor pool. A larger pool of software industry workers within commuting distance allows for easier expansion of a firm through the poaching of employees of other firms. In the absence of any additional flows of employees into the area, this is a zero-sum game in terms of "bodies." However, the overall effect on firm productivity and growth could be positive if employee-employer match quality improves by the turnover and if a larger pool of employees increases the propensity of software engineers and programmers to move into this area. The labor channel can also operate through the founding of start-up firms by employees departing their current employer and creating another firm. These firms are often located very close to their "parent" firms, sometimes in the same parking lot. Thus, labor pool effects could have a very localized component and also be present at commuting distances (though attenuated). As with knowledge spillovers, there is the likelihood that these effects have both winners and losers; some successful firms may feed off the labor pool of their neighbors. In fact, as one industry executive confided to us, programmers employed by different firms compare their work conditions and terms of employment when in social contact and are ready to switch employment when their current employer is not competitive.[3]

[2]Though extremely unlikely, spillovers might conceivably even be negative in aggregate if they lead to free-riding. For example, firms may reduce experimentation with new ideas, and hope to piggy back on ideas developed by other firms, thus possibly leading to a reduction of available knowledge.

[3]Some Silicon Valley firms run company buses from San Francisco to their facilities for the purpose of easing commuting. Company buses also reduce, incidentally or intentionally, employee interaction across firms (relative to car pooling, public transport, or other non-company transport), suggesting that such interaction is not valued by firm.

Both knowledge spillovers and labor pool effects are likely to lead to a positive association between agglomeration and spatial persistence in the software industry. However, the effects are expected to differ subtly on the sources of this persistence, i.e., whether it is driven by the growth of the existing establishments in a location or by the entry of new establishments and their survival.[4] Labor market effects may attract entry through the availability of labor and lead to the retention of jobs in a locale even without improving the survival prospects of the new entrants or the success of the existing firms. Labor markets effects can also improve firm growth and survival if thicker labor markets improve employee-firm matching and reduce ongoing costs of recruiting. But in the absence of any observed entry effects, increased growth and survival rates are unlikely to be driven by labor market effects. In contrast, knowledge spillovers are more likely to lead to gains for the existing firms and better survival prospects for the entrants. Of course, by increasing growth and survival prospects, knowledge spillovers would likely also lead to increased entry. However, in the absence of any observed growth and survival effects, increased entry propensity is unlikely to be driven by the knowledge spillovers.

Though this discussion does not constitute a formal model, the comparative statics we describe can be illustrated diagrammatically. Figure 1, panel A, illustrates how entry and firm performance outcomes differ across two locations, one with low knowledge spillovers, the other with high knowledge spillovers. Firms are distinguished by a summary scalar measure of their relative (or ordinal) performance, which can be variably referred to as productivity rank, or business model riskiness, or frailty. Relative performance is plotted on the horizontal axis, with higher relative performance firms closer to the origin. The vertical axis plots a measure of firm value as an ongoing prospect, which can be referred to by any index of cardinal performance such as profitability, growth, or survival probability. The downward sloping lines are the relationship between firm value and relative firm performance for low and high spillover locations. By the definition of relative performance, both lines have a negative slope, but the line for the high knowledge spillovers location is above the line for the low knowledge spillover location. The premise is that high knowledge spillovers increase the performance of all firms (though not necessarily by the same amount; we have drawn the lines with different slopes suggesting a bigger positive effect for high relative performance firms). A firm that enters a location must expend an entry (or setup) cost. For simplicity, assume that this is the same for all firms in a location, and plotted by the dashed horizontal line at FC. The marginal entrant in the low and high

[4] We note parenthetically that employment growth and exit are qualitatively different in that exit is a "tail event." Thus, an increase in the variance of outcomes may not affect mean firm employment growth, while affecting survival probabilities.

spillover location is $\hat{E}_{low}$ and $\hat{E}_{high}$, respectively. High spillover locations attract more firms. Moreover, the average cardinal performance of these firms is higher, i.e., firms in these locations should grow faster and be less likely to fail on average, despite including entrants that are more "marginal" in a relative sense.

Panel B of Figure 1 illustrates how entry and firm performance outcomes differ across two locations, one with low labor market agglomeration economies, the other with high labor market agglomeration economies. Agglomeration effects arising from the labor channel reduce the cost of entry into a location, from FC_{low} to FC_{high}. They can also reduce ongoing recruiting costs and thus shift up the post-entry value of establishments (profits, growth, survival). This second effect is expected to be smaller, but whether or not this is the case, what is important is that the labor channel affects both set-up costs and also future performance. As can be seen from Panel B of Figure 1, average (cardinal) performance of firms in these markets may be lower, even though they attract more establishments. In fact, we believe that it is likely lower, since the entry effect is expected to predominate, but whether or not this is the case depends on the relative size of the shifts of the two lines.[5]

These two figures are both static representations, and can be thought of as describing entry outcomes in a single period under a deterministic setting. Naturally, spatial agglomeration will persist over time through differential entry, growth, and survival probabilities across locations with different spillover levels. But persistence will not be perfect, as random variation in firm outcomes will weaken the initial advantage of some locations and strengthen that of others. Spatial persistence may be particularly high if the source of agglomeration economies is the labor pool channel. When a firm receives a negative shock and lays off workers, it creates a pool of local job seekers and reduces the recruiting costs of co-located firms, facilitating their expansion. This tends to stabilize employment at the location.

The estimation of the spatial spillover effects needs to take into consideration that there are a number of other relevant economic factors that co-move with entry propensity, employment growth, and exit of software publishers. Obviously economic conditions are likely to be important, as are the national prospects of the industry. But local factors are also important. Some parts of the state are growing and becoming more attractive places to live; firms are more likely to locate into those places and be better able to attract workers there. Though there are county-level data that can help account for these effects at a low level of

[5] Upon reflection, one can discern that heterogeneous entry costs do not alter any of this discussion as long as firms with high post-entry value enter" ahead" of firms with low post-entry value. A more substantial modification of this framework, under which firms have the same post-entry performance but differ only in the entry costs (represented by a flat post-entry cardinal performance curve and upward sloping entry cost curve), results in some meaningful changes in the comparative statics, but cannot be reconciled with the empirical results.

spatial granularity, the identification of localized spillovers requires controls of higher spatial granularity. Consider the following illustrative example. A location that was hitherto farmland might become developed and office parks may be built. These office parks will be filled by a number of "white collar" employers, regardless of the presence or absence of spillovers. Thus, a positive spatial association at very small distances may be an artifact of land development patterns (or, equivalently, the abandonment of commercial land that has become less desirable economically). We are cognizant of the importance of these effects, which are pertinent to studies like ours that aim to identify spillovers at distance of a mile from spillovers over a larger distance. We account for overall localized economic activity by defining, based on the empirically observed co-location pattern, a set of industries which use facilities similar to those used by software publishers. The employment in these industries is used as a control in a number of ways as explained further below.

We discuss the estimation procedures and results for each type of analysis (employment growth, entry frequencies, and exit rates) in separate sections. However, all econometric models utilize a common underlying primary dataset and share a number of independent variables. For that reason, we first provide an overview of the data and the construction of the common set of independent variables.

3 The Dataset and Spatial Patterns

3.1 Data and Variables

The establishment data used in this study are from the Quarterly Census of Employment and Wages (QCEW) compiled by the Texas Workforce Commission. This dataset provides establishment-specific monthly employment and quarterly total wages reported for all firms as required under the Texas Unemployment Insurance (UI) program. Each record includes the specific location (address and latitude/longitude) of the establishment, business start-up date (the date on which UI liability begins), and the relevant six-digit North American Industrial Classification System (NAICS) code. Separate establishments (branches or franchises) of the same firm are identified and reported in separate records. This highly detailed panel dataset is comprised of observations from the first quarter of 2000 through the fourth quarter of 2006 and allows us to map the locations and calculate distances between any two establishments.[6]

After restricting the analysis to the software publishing industry (NAICS code 511210), the sample has more than 15,900 observations, with 957 establishments corresponding to 877 unique firms (the vast

[6]The authors obtained these data under an agreement of confidentiality and, therefore, disclosure of the actual data is subject to certain restrictions. While the data is restricted to the State of Texas, it should be borne in mind that Texas is a large economy. Over the period of analysis, the Texas economy was the second largest among US states, and would rank in 2008 as the 15th largest economy in the world. Moreover, Texas is geographically both extensive and diverse.

majority of firms are single-establishment enterprises, and thus establishments will be used inter-changeably with firms). Average firm size is relatively small, around 35 employees.[7] The number of software publishing firms decreased from a high of 648 in Q4:2002 and Q1:2003 to a low of 581 in Q4:2006 (there were 526 firms in Q2:2000).[8] However, employment in the industry increased from 16,600 to 21,000 over these seven years, implying an increase in average establishment size. Key features of the data are provided in Table 1.

We eschew political or administrative geographical partitions in our definitions of locations and their neighborhoods, because these partitions tend to be large and often irregularly shaped. Instead, our locations consist of circles and their neighborhoods of concentric rings. The choice of ring radii reflects our interest in estimating the extent of the localization externality. Most Texas counties and many MSAs would fit into a ring with a 50 mile diameter.[9] So a 25 mile radius, which is the radius of the outer most ring, approximates a county-level analysis. As noted above, earlier researchers have concluded that the geographic extent of knowledge spillovers is quite limited. Rosenthal and Strange (2003) conclude that localization externalities largely dissipate within a mile. A two mile diameter also captures a majority of most central business districts. Thus a 1 mile radius forms our inner circle. Our choice of the further subdivision middle ring at distance thresholds of 5 and 10 miles reflects a "windshield" observation of the urban landscape in Texas cities. That is, commercial (high rise) clusters are distributed discretely across the urban landscape. The intermediate rings of 1-5 and 5-10 miles capture the concentration of similar firms in adjacent commercial clusters or in the effect of possibly isolated firms in the space between clusters, but do not capture all firms within the city or urban region. To summarize, we define locations to be 1 mile radius circles, which are surrounded by rings whose outer boundaries are at 5, 10, and 25 miles away from the center of the location.

One concentration measure we consider is the number of firms in a given location. This is simply a count variable of the number of other (or rival) software establishments. We do not normalize by area size; however, the variable is effectively normalized since it is measured in each observation for each firm in terms of identical areas as defined by the concentric rings. The other measure of industry concentration that we use is the number of software employees in rival firms in a given location. Construction of this variable is similar to the number of rival firms variable described above. Within one mile of an existing software firm, there are, on average, approximately 10 other software publishing firms employing 394 employees. At distances

[7] Acs et al. (1994) find that smaller firms may benefit disproportionately from knowledge spillovers.

[8] There were 14 software publishing establishments with PO Box as the official address, and for which physical location could not be ascertained. These are not included in the above totals or in the subsequent analysis.

[9] The majority of Texas counties are approximately square. The average shortest distance from the center point in the county to the county line is approximately 22 miles.

between 1 and 5 miles, there are an additional 48 software firms employing 1,857 employees, a substantial drop off in the density of software firms (recall that area is proportional to the square of distance). The next 5 and the following 15 miles contain 46 and 72 software firms employing 1,810 and 2,525 employees, respectively. These correspond to ever larger drop-offs in density. There is a similarly steep drop-off over these distances in the density of firms in the white collar service industries that form our control (and which we discuss in detail in section 3.3). Control industry employment goes from approximately 16,500 within a mile of a software establishment, to 155,000 in the next 4 miles, 314,000 in the next 5, and 303,000 in the next 15, reflecting the localized "strip mall" nature of commercial development in Texas and the average size of Texas cities.[10] Note, however, that in small scales (up to 10 miles) the employment of software firms seems to drop off faster than that of the control group of firms.

Considering firm characteristics, we include a measure of the firm's exposure to university R&D funding. This variable captures the possibility that knowledge spillovers are available from research universities.[11] This variable is measured as total federally funded research expenditures at the university located closest to the firm, using the main address for the university campus. In order to introduce distance decay in the university R&D expenditures, we deflate total R&D expenditures by distance in miles. We distinguish between research universities and colleges/junior colleges. Colleges may be smaller 4-year or 2-year degree-granting institutions. College funding is treated similarly to university funding. We include junior college funding since previous studies have found this variable to be more important than university research funding in explaining high-tech firm location decisions (Abramovsky, 2007).

We proxy the quality of the labor pool and the ability of firms to attract skilled employees by including a measure of local recreational and cultural amenities. As Woodward et al. (2006) suggest, cultural and natural amenities are important to industrial attraction and skilled workforce retention. A well educated and tech-savvy workforce demands a high and diverse level of urban amenities. To measure the relative local

[10] One might be surprised by the large employment counts, which amount to nearly 800,000 employees within 25 miles of the average establishment. The reason is that these averages are taken at the establishment-cross-quarter level. With most software firms being located in the metropolitan areas (especially Dallas and Austin), the typical urban density dominates the data. Note that in Dallas county there are approximately 1.3 million employees in the control industries (about 1.4 million total employment), the Houston numbers are similar, while those for Austin (Travis county) are 540,000 and 560,000, respectively. Observe the that the bulk of employment in those areas is in the control industries. For the entire state of Texas, approximately nine-tenths of the employees are in the control industries, something that we discussion in section 3.3.

[11] Note that we assume that knowledge spillovers from universities are proportional to the level of research conducted at the institution. 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. Data on annual University R&D expenditures were obtained from the National Science Foundation (NSF). The annual NSF data actually span two calendar years, since the federal fiscal year begins in October. In order to convert these annual R&D expenditures into quarterly data, we use a fourth of a given year's total for each of Quarters 1-3, and a fourth of the following year's total for Quarter 4. We aggregate total federal awards by all granting agencies, i.e., DoE, EPA, DoD, by geographically distinct institution, i.e., system campuses are scored geographically separately.

presence of these amenities, we compute the share of county employment in NAICS 71, Arts, Entertainment, and Recreation, and NAICS 721110 (hotels and motels), 722110 (full service restaurants), and 722410 (drinking places, alcoholic beverages) as reported in the QCEW data set. A similar measure has been used by De Silva and McComb (2012).

To account for factor costs, we use the average quarterly payroll of high-tech industries in the county.[12] The county unemployment rate for the final month in each quarter, as reported by the Texas Workforce Commission, is included to provide an indication of the overall economic conditions in the local county. Summary statistics for these variables are given in Table 1. Note that in the bulk of the regression analysis we exclude firms that have single employee throughout our sample.[13]

3.2 Some Facts about Software Firms and their Spatial Distribution

Before proceeding to formal econometric analysis, it is instructive to describe some relevant key characteristics of software firms and their spatial agglomeration. As suggested by the summary statistics in the preceding section, these firms are rather numerous, to the surprise of the authors. Software firms are not completely undifferentiated as they tend to specialize in specific market segments. An important question is whether software firms in the same market segment tend to concentrate in particular locations. This is important because, if true, it would suggest that any human capital spillovers may be specific rather than general. It would also suggest that co-located entry might be driven to some extent by employees of a firm who launch a spin-off in a different office of the same building or in the building next door. Moreover, the interpretation of some of the findings might be affected. In particular, we assume that software firms do not compete with each other in the local market, and therefore that local demand effects and supply considerations have no bearing on the spatial association of outcomes. This is a reasonable assumption. If, however, software firms that are in the same market segment are all co-located (for whatever reason), then national swings in the demand and supply of these products would induce a spatial correlation in outcomes. For example, if the demand for Voice Over Internet software were to increase, and firms were to grow and enter in that market, then spatial specialization would result in a positive spatial association between firm entry and growth.[14] More importantly for the econometric analysis in this study, if establishments in some

[12]We considered the inclusion of the undeveloped land price, but that measure was not available even at the county level, and probably a poor proxy for the software industry given the other controls.

[13]There are only eight such firms, but five of them are new entrants. They are only included in the entry analysis, since they may inform the decision of where to locate.

[14]If indeed observed, such a co-location of firms in the same market segment might itself indicate some spatial interaction that is market-segment specific.

business segments tend to be spatially concentrated while establishments in other segments tend not to be spatially concentrated, then the an increase in the relative market size of the former business segments would result in a correlation between growth, entry, and concentration.

Table 2 provides information on the location of individual firms that can help assess this localization, albeit informally. We randomly select five firms in separate cities, labeled A, B, C, D, and E. Then, for each of these five firms, we list all the other software publishing establishments located less than five miles away. We indicate whether firms share the same building, their primary business function, their average employment level, and the years they were operating.[15] The firm and city information needs to be anonymized to conform to the requirements of obtaining the data, but the information in the table is still instructive. There are seven pairs of firms that located in the same address. Looking first at these pairs of co-located firms, we see that in two cases a building contains a large firm along with a much smaller one. But the smaller firm does not appear to be an off-shoot of the larger firm; for both of these pairs, the firms are in different software categories. Overall, taking all seven pairs of co-located firms, there is one pair of similarly sized firms that belong in a similar market segment, and one pair that seems to have generic labels (and where similarity cannot be fully assessed). Most of the other pairs are clearly discordant in terms of market segment. To expand our sample of co-located firms, we also looked for such firms in selected locations other than the five cities reported in Table 2. This more than doubled the sample of co-located firms, but resulted in the same qualitative pattern.[16] There are now proportionately slightly more pairs of large and small co-located firms than in Table 2, but all pairs except one consist of firms in different business segments (these firm pairs are not reported in the interest of conserving space, but can be readily be made available by the authors). Though the sample is small, and we can not make any formal inferences, it appears that there might possibly be some minimal agglomeration of firms in the same line of business at the building level, but such agglomeration, if indeed present, is very small. Moreover, there seems to be no evidence of co-located firms being spin-offs of existing firms entering the same business segments as the originating firm, though some collocated firms may be spin-offs venturing into different business segments.[17]

Let us now turn to the examination of the firms in each of the five mile radius circles reported in

[15] The primary business function was obtained by searching for each firm on the Internet. This labor intensive procedure cannot be replicated for every single facility in our dataset. For this reason, this evidence is presented for a small sample and the analysis is somewhat informal by necessity. For city B, there were approximately 25 firms within 5 miles. Thus, we report firms within 1 mile of selected location.

[16] In total, we examined approximately 40 of the 181 co-located firms in the entire state of Texas.

[17] Buenstorf and Klepper (2009) have demonstrated in their study of the Akron tyre cluster that organization replication can generate a spatial pattern akin to that arising from agglomeration economies. This seems to not be the case in the software industry.

Table 2. An examination of the descriptions of the business segments to which the firms belong to reveals some categories that are more common than others (e.g., "computer software" or the "business" descriptor) perhaps because of the more generic nature of these labels. In city A, the firm in the "bull's eye" is one such generic producer of both software and services. Two other firms in its area list computer software in their business category, but both omit the provision of services. In city B, the selected firm is in Electronic Trading Software, an uncommon category. A nearby firm seems to be in a related market segment, but the others are very distinct. The firm in city C is a service provider; 3 of the other 15 software firms in each area emphasize the provision of services, but so do some firms in the other listed cities. In city D, a firm focusing on "Business solutions" is located close to another two firms in related market segments. Finally, in city E, a firm specializing in Entertainment software is located close to another firm in the same business; but that firm was tiny and lasted only three months (clearly a failed attempt to launch a business). We repeated the exercise for a random firm located in the Dallas metropolitan area, where the number of firms within the five mile radius is too large to report in this table (we can make this separately available upon request). The similarity of firms within five miles of the selected firm was very small, even though that firm specializes in computer game software which is not a "rare" category. Only one of 31 firms located within five miles was a computer game developer. Moreover, the list of firms does not seem dominated by any particular software category. Though there are a number of categories with more than one firm, that would be expected simply by chance.

To summarize, while the totality of the evidence is against the notion that there is strong spatial specialization, a small amount of such specialization cannot be ruled out. It appears that a firm is slightly more likely to be surrounded by firms in the same business segment as itself, but the large majority of software firms in the same area are engaged in a different part of the software publishing business. It is thus unlikely that co-location is driven by spin-offs or by human capital that is specific to a market-segment.

3.3 Location Definition and Sampling

A key feature of our analysis is to identify the geographic scale at which agglomeration is relevant, which necessitates defining locations at the sub-county level.[18] In principle, this would be accomplished by looking at county sub-divisions, such as census blocks, but these are of variable geographic extent, irregular shape, and are constrained by county boundaries. Ideally, we would like locations to be equally sized and defined

[18] Some of the variables discussed in sub-section 3.1 are by necessity defined at the county level.

in a purely geographical manner, while ensuring that they are not overlapping. Locations defined in this manner would be more easily used as units of analysis for estimating the spatial persistence of software industry employment and also for estimating the entry rates of software establishments.

A desirable feature for the definition of these locations is that they only include plausible destinations for software establishments. The inclusion of destinations with essentially zero probability of containing a software establishment is problematic in two ways. First, it would not be consistent with specifications which implicitly assume a positive expected number of software firms, i.e., it would require the use of models where for most locations an outcome of no entry happens with probability 1 regardless of covariate values. Second, it would result in a plethora of observations/locations, only a tiny fraction of which would experience software firm presence over our sample period. In defining locations, we adopt the principle that human capital requirements can always be potentially met, at some cost, at any location either because required employees are already available for hiring locally or because they could be induced to move into the area. However, the locations of firms are restricted by zoning laws, by the availability of suitable building stock, and by the presence of complementary infrastructure (e.g., roads, utilities). We posit that the likelihood that a software firm locates in a location where these conditions are not satisfied is zero.

Our approach to identifying potential locations is as follows. More than 90 percent of software establishments share a building or address with other non-software establishments. We refer to the industries that these other establishments belong to as "control industries" or (less imaginatively) as "other industries." There about 700 industries on this list (at the 6-digit level) which constitute approximately half of all industries in the state of Texas (and most of the employment).[19] We take it as evidence that any location that contains an establishment of these other control industries is a location where a software establishment could potentially be situated (though we recognize in our estimation approach that the probability of doing so is not the same across all such locations). In other words, we assume that the physical infrastructure embodied in a building is fungible across these industries. Clearly, this is not literally true. For example, if we observe that a software firm shares a building with an advertising agency, it does not mean that every building that houses an advertising agency would potentially house a software firm. But most such buildings would or (equally importantly) most such buildings would at least be in close proximity to other

[19]The reason for the large number of industries that share facilities with software firms is that many of industrial and agricultural firms also have office operations located separately from the production facilities. These offices are sometimes in the same building as software establishments. A weight scheme, which we describe at the end of this section, mitigates any associated issues when these industries are used to construct control variables.

buildings that would, and thus provide a reasonable starting point for identifying plausible locations which could potentially house software firms. Moreover, it is extremely unlikely that a software firm would locate in an area that does not contain at least one establishment on this long list of control industries.[20]

We then take all establishments in the control industries that have ever operated during our sample and retain their co-ordinates (dropping duplicates). We use these co-ordinates to obtain one-mile radius (non-overlapping) locations. Note that a location is in our sample even if it contained a control industry facility for only part of our sample, i.e., when these locations are used in a panel, as in the entry analysis, they yield a balanced panel of locations. For example, a location that contained no control firms in the first two years of the sample will be in the sample during those first two years as a potential entry location, thus permitting a software publisher to be the first entrant in that location. In particular, there are 9,936,068 observations of related industry establishments, which correspond to 580,375 unique establishment locations (including establishments that entered or exited during our sample period).[21] We sorted these establishments by longitude (from west to east) and went sequentially through them dropping any establishment that was closer than 2 miles to a previously selected establishment. We obtain a final sample of 9,299 establishments, which form the center point of circle locations with a radius of one mile. These locations cover about 11 percent of the area of Texas, and are shown in Figure 2, Panel A (for the entire state of Texas) and Panel B for the Dallas-Fort Worth area.[22] When used in a panel over our 28 quarter sample period, this results in 251,073 observations (27 quarters are used in the regressions, since the first quarter is dropped because of one quarter lagging of some variables). In Table 1 Panel B we report the summary statistics for these non-overlapping locations. We note that nearly all software establishments are less than a mile from a control industry firm, and thus these control industries effectively map out the large majority of potential locations.[23] However, after we drop locations to eliminate overlap, some gaps in space are created, and thus the proportion of software firms that are outside the final set of circle locations rises to 28%. As will become apparent from the analysis that follows, this does not create any issues beyond the loss of these

[20] Our approach of identifying locations in this manner also controls for potential zoning regulations, without the need to actually obtain the geographic extent of areas zoned for commercial development. If two industries are co-located then there are probably few (if any) zoning regulations that permit the presence of one but not the other. The presence of zoning and other legal or non-legal impediments to co-location should be reflected in the relative frequency of co-location. This is also discussed in more detail below.

[21] There were 11,791 establishments with Post Office addresses. The location of almost all of those was obtained via batchgeo.com, and are thus utilized for the purpose of controlling for localized economic activity and infrastructure.

[22] The final list of establishments depends on how the initial sort is made, as the maximal set of establishment locations with mutual distances that exceed 2 miles is not unique. For example, when the initial sort is on the basis of latitude (from south to north) the number of establishments we obtain differs somewhat. However, the covered area differs very little, as we pick up most of Texas that contains commercial activity.

[23] There are only 14 establishments for which the nearest control firm is more than 1 mile away. Most of those are one person operations (likely a free-lancer working from home).

observations from the sample.[24]

The control industries are not only used in the definition of locations, but also as a control of the baseline propensity of software establishments to be situated (or enter) there. The necessity for such controls follows from the observation that those locations have some infrastructure that makes software establishment location possible, but differ in the extent to which they posses that infrastructure. Thus, the propensity of software firms to locate in these locations differs in the extent of their development. Locations that contain a large number of control industry employees or firms would, all things equal, contain more software establishments. Less developed locations, with a smaller number of control industry employees or firms are, all else equal, likely to contain fewer software establishments. Therefore, the number of control industry employees is used to account for the baseline presence of software establishments or employees. Not controlling for this baseline might generate spurious spatial persistence in employment or clustering of entering firms, since an area that is becoming more developed attracts more firms, including more software firms. We use two different ways to account for this baseline. The first way is to use the number of employees in the other industries as a control variable. The second way is to use as a control variable the weighted number of employees, with weights obtained from how frequently establishments of each industry are co-located with software establishments. In this second approach, the employees of each industry have an industry specific weight, which is the fraction of facilities in that industry that are collocated with software facilities. We find the second approach more appealing, and report the results using the weighted controls in the paper. However, the results using the unweighted number of employees are qualitatively similar to those using the weighted measures (and are available from the authors).

4 Econometric Analysis and Results

4.1 Spatial Persistence in Software Industry Employment

The software industry is a very dynamic one, with a large turn-over rate of establishments and substantial changes in the scale of establishments over time. In particular, approximately half of the establishments that were in operation at the start of 2000 had exited by the end of our sample in 2006, while many of the establishments that did not exit experienced large drops in employment. All in all, from the 16,645

[24]The only way to avoid gaps that potentially contain software firms is to divide Texas into a fine grid, creating squares of (say) one mile width, and dropping those that do not contain any control establishment. However, these squares would in general not have a control establishment at their center, and would cover a much larger portion of the state. Moreover, it would be geometrically challenging to define neighborhoods around square locations that contain all points that are no more than a specified distance from the edge of the location.

jobs that were initially in our dataset, only 7,015 (or about 41 percent) persisted in the same establishment until the end of the period.[25] We emphasize that persistence of jobs is not the same as worker turnover; a worker who leaves an establishment and gets replaced by another worker at that same establishment during our sample period registers as a retention of that job by that establishment. During the same time frame, 372 new establishments entered (almost all in different buildings than the exiting establishments). The jobs created by these entrants and the jobs added by growing incumbent establishments raised total employment in the industry to about 21,000 (an increase of 24%). In other words, only one third of the jobs at the end of the period were jobs that existed in the same establishment at the start of the period. In principle, given that this industry does not rely on specialized infrastructure (and has non-localized demand for its product), and given the entrants typically choose different addresses than incumbent or exiting firms, there is a potential for the spatial distribution of this industry to be completely transformed.

However, this turns out not to be the case. At the macro level, a quick way to assess whether spatial concentration has increased is to investigate whether the share of employment in, say, the top five counties has increased over this period. The identity of the top five counties has remained the same (Dallas, Travis, Harris, Collin, and Bexar). The number of software publishing employees in those five counties has increased at approximately the same rate as in entire state of Texas, marginally raising their combined share from about 89% at the start of the period to about 90% at the end. It is worthwhile to point out that, while these figures indicate that concentration is (slightly) increasing at the county level, they do not provide any direct evidence about concentration at the 1-mile radius level.

Some evidence at the micro level can be provided by examining a few representative areas pictorially. One cannot easily plot employment into space, but we can plot establishments. In Figures 3 and 4, we have plotted the software publishers and the control firms in Austin and north Dallas areas for the first and last quarter of our sample. The distribution of control firms indicates the areas near which software publishers could be located. The first thing to note is that software publishers are not uniformly distributed in this space, but are rather concentrated in sub-areas. The second thing to note is that the areas of concentration remain stable even if there is entry and exit. An area with prior concentration of software publishers seems to retain that concentration to the end of our sample period. A final observation relates to newly developed locations. It is clear that some areas that were not commercially developed in the second quarter of 2000

[25] This percentage is equal to $(\sum_i min\{emp_{i,1}, emp_{i,T}\})/(\sum_i emp_{i,1})$ where $emp_{i,1}$ is an establishment's employment in the first quarter of our sample (and 0 if the establishment entered at a later date) and $emp_{i,T}$ is an establishment's employment at the last quarter of our sample (and 0 if it exited by that quarter). Note that by construction this ratio cannot exceed unity.

became developed by the end of 2006. However, these areas do not seem to have also spawned a cluster of software firms. Moreover, isolated software firm entry into these newly developed areas seems to be rare.

Some additional evidence at the micro level is obtained by computing the number of software employees in each of the 1-mile radius locations we have defined and then see how that number changed over our sample period. The top ten 1-mile radius locations in terms of initial employment contained 61% of software employees in our initial quarter. The corresponding percentage at the end of our sample period is 65%, an increase in concentration despite an increase in the number of establishments and the number of employees. However, there has been some churning of the top 10 locations. The percentage of software employees employed in the end of the sample period in the top 10 locations measured by employment at the start of the sample period is 48%, a decline from the initial level, but still a remarkable persistence in such a dynamic industry. Removing from all calculations one outlier location which jumped to first place from outside the top 10 list due to the entry of a large facility, the percentage increases to 57%, essentially unchanged from the initial level of 61% despite the dramatic changes in the landscape of the industry!

A more systematic and formal analysis would consider all locations with positive employment at the start of the sample period, and account for other changes in these locations that would be expected to affect end-of-period employment. The conceptual exercise we want to perform is the following. Suppose we increase the software jobs (or employment) in a particular location at the start of the period by one percent (this could happen by, say, increasing the order flow and product demand for the establishments in that area, or by swapping the establishments in that area with other establishments that are slightly bigger). What would then be the percentage increase in the jobs (or employment) in that location at the end of the sample period?[26] Jobs, to be clear, are not the same as employees. In all discussions here, an employee leaving an establishment only to be replaced by another employee in that same establishment has no effect on the jobs in that establishment.

Of course, we do not observe software jobs being exogenously created in a location and measure their impact on that location's jobs at the end of the sample period. Rather, we observe locations that differ in the level of initial software employment and other characteristics. Our attempt, using locational controls based on industries that are typically co-located, along with the special feature of the software industry that it has minimal infrastructure demands and a non-physical product with a national market, ensures that the

[26]The same hypothetical question could be expressed in number of employees rather than percentages, in which case it would correspond to analysis using the number of employees rather than their log. See discussion on this alternative analysis below.

initial variation in employment comes close to being exogenous in the statistical sense.[27]

To better understand the content of the question we investigate, we observe that end-of-period jobs in a location are a function of jobs lost and accrued, the latter set consisting of accrual to entering firms and accrual to incumbent firms. For the moment, suppose that every job (or position) in all establishments has the same probability of being lost and that job accrual is proportional to the initial number of jobs in a location. Then, the elasticity of final period employment with respect to initial employment would be equal to 1. However, such an extreme degree of persistence is unlikely. Entrants, for example, are more likely to locate in places with more existing software jobs (conditional on overall economic activity), but not proportionately more likely. Similarly, establishment job growth is not proportional to establishment size (as shown by the voluminous literature on Gritbat's law), and hence final employment in a location is not expected to be proportional to initial employment in that location. As a result, there would be some shift in the landscape of software industry activity over time: some areas with high concentration would "revert to the mean" closer to a level of software industry activity that is in proportion to overall economic activity in that location; other areas with limited activity might exhibit higher concentration for idiosyncratic reasons. Spatial persistence would be higher than this benchmark if establishments located in areas with a high concentration of software firms grew systematically faster relative to establishments located in areas with low concentration of software firms. In sum, the degree of spatial persistence of employment would balance these factors. In the absence of agglomeration economies, we would expect spatial persistence be driven solely from the inertia of jobs at the establishment level.[28] With agglomeration economies, we would expect it to higher than this value.

We now describe the econometric framework through which we investigate the extent of spatial persistence. Our unit of econometric analysis is the one mile radius circle locations described in section 3.3. For each of these locations, we estimate the end-of-period employment in the software industry as a function of initial employment and other initial conditions. We use as initial conditions the number of software firms located inside the location at the start of the sample period, the number of software firms in concentric 1-5, 5-10 and 10-25 mile disks and the associated number of employees of these firms. Though the 1-mile radius locations are not overlapping, the surrounding rings are, as shown in Figure 5, where Panel A shows the locations in the Dallas area, and Panel B shows the surrounding rings in two of those locations.

[27]It is not required that initial variation in employment be random, but only that it is uncorrelated with unobserved shocks to end-of-sample period employment.

[28]This assumes that locations are "small" as a fraction of the industry employment.

The overlap of the rings does not raise any econometric issues. County effects for the five counties with major employment in this industry are included in some specifications.[29] For these regressions, the effect of initial conditions is identified from the within-county distribution of software publishing firms. Clearly, some locations experience more rapid development than others. For reasons explained above, we include the (weighted) number of employees in the set of control industries in the initial and final periods as explanatory variables in a number of specifications. In addition to these exogenous variables, we sometimes also include the final period number of firms and employees in the concentric rings surrounding a location, which are to some extent co-determined (though the stronger influence probably goes from the larger outer rings to the center). More formally, our most general regression model is given by

$$\ln(emp_{l,T}+1) = D'_{l,t=1}\delta_0 + R'_{l,t=1}\rho_0 + C'_{c_l,t=1}\varphi_0 + D'_{l,T-1}\delta_1 + R'_{l,T-1}\rho_1 + C'_{c_l,T-1}\varphi_1 + \eta_l \quad (1)$$

where $emp_{l,T}$ is the location l's log quarterly software employment in the final quarter, T (Q4:2006). The value of 1 is added prior to taking the log of employment or the number of firms, as done elsewhere in the paper. These variables have long right tails and can sometimes take the value of zero. Indeed, in the subsequent analysis on entry, their typical value is zero. Moreover, many of the explanatory factors are expected to act synergistically, rather than in a purely additive fashion. Thus, an analysis using employment (rather than its log) as the dependent variable is less appropriate and results in larger standard errors, though the main conclusion with regards to employment persistence is robust (these results are available from the authors). The vectors $D_{l,t=1}$ and $D_{l,T-1}$ represent initial and final period (minus a lag) density variables for location l, respectively. Similarly we include initial and final period (minus a lag) location variables ($R_{l,t=1}$ and $R_{l,T-1}$), and variables specific to the county c_l where this location is situated ($C_{c_l,t=1}$ and $C_{c_l,T-1}$). Only locations with positive initial software industry employment are included in this regression.[30] Locations with zero final employment are included to avoid truncation bias and the regression is estimated both via OLS with robust standard errors and via Tobit.

Table 3 presents the linear regression results with robust standard errors, generally moving from the simplest to the more complicated specifications, while Table 4 presents the results with the exact same

[29]The top 5 counties for which we include dummy variables are Dallas, Travis, Harris, Collin and Bexar. Using country fixed effects for all counties effectively "dummies out" most locations situated in those counties. For sensitivity analysis, we also estimated those models, resulting in no major changes in the findings.

[30]Including in the regressions all locations yields estimates that suggest even stronger spatial persistence, in fact elasticities that exceed unity (implying that an area builds on its initial advantage). However, these results are due to a strong bias introduced by the inclusion of those observations. Observe that these observations would be at the origin (log of initial employment plus 1 and log of final employment plus 1 are both zero). The regression line has a positive intercept, so adding a mass of observations at (0,0) pushes the intercept down and the slope up.

specification estimated via a Tobit model. Estimating the Tobit specification is important, despite its relative lack of robustness, since one quarter of the observations are censored. This percentage seems a bit high given the large persistence in employment, but about 70 percent of the locations have only one firm, and these locations form the bulk of those that are censored (i.e., censoring essentially implies that the sole establishment in that area has exited and has not been replaced). When viewed in this way, and given that half the incumbent establishments fail, the censoring fraction appears low, as it implies than in many locations with only one establishment at least one other establishment was attracted prior to the exit of the incumbent establishment. This is remarkable given that the locations with software firm activity are but a very small fraction of possible locations that software firms can choose to locate, revealing the propensity of entrants to locate in close proximity to incumbents (a point that will formally shown in section 4.2).

We discuss the results of Tables 3 and 4 together. In the simplest specification (model 1), no covariates except for differential intercepts for major counties are used. The point estimates suggest than a one percent increase in initial employment translates into 0.78-0.86 percent increase in final period employment, which indicates very large spatial persistence in line with the informal evidence described earlier. The larger of the two estimates in all specifications corresponds to the Tobit results, something that is expected from the standard graphical intuition of censoring in a single variable regression model. Recall that the parameter estimate in the Tobit is the coefficient of the initial employment on the latent variable. More directly comparable to the OLS parameter estimates are the marginal effects of a regressor on the expected value of the dependent variable in a Tobit model. This is equal to the sum over all observations of the probability that an observation is uncensored, times the parameter estimate, divided by the number of observations. We have computed these marginal effects for the key variable of interest, the log of the initial number of employees in a location, and report it to all Tobit tables. By definition, the marginal effects are smaller than the coefficient value and thus this brings the Tobit and OLS estimates closer in line with each other, though the former continue to be somewhat larger (the value for column 1 is an elasticity of 84 percent).

The addition of software employment in close and moderate proximity and other location characteristics (model 2) has no material impact on the elasticity estimate. Interestingly, while location characteristics affect (jointly) the employment level, employment at any distance up to 25 miles has no such effect. Controlling for localized employment and growth in the control industries (model 3) somewhat reduces the elasticity estimate. Adding current location conditions to the regression (model 4) shows that final employment in a location increases with initial proximate employment but decreases in final proximate employment (lagged by one quarter). Because employment in a mile-radius location is typically a small fraction of the employment in the surrounding 25 radius ring, we posit any causal effects go mostly from the surrounding area to the inner circle. Thus, locations in moderate proximity to an existing software cluster grow faster, but those in close proximity to a growing cluster grow slower. Knowledge and other productivity enhancing spillovers are unlikely to yield this pattern (software activity, whether initial or final period, should increase employment in a location), but can easily be rationalized with a labor pool thesis. An area in proximity to many software employees can provide a ready supply of workers who can easily be attracted to establishments in a location whereas a growing area in proximity can siphon those workers away.

The last two models (5 and 6) augment these specifications by adding the number of establishments at various distance thresholds as explanatory variables. In these regressions, agglomeration effects are decomposed to those arising from an increase in the number of facilities (holding employment constant) and those arising from an an increase in employment (holding the number of facilities constant). Both appear significant and of approximately equal magnitude, though the coefficient on the number of firms is less precisely estimated. Remarkably, a replication of the facilities in a location (i.e., doubling the number of facilities active at the start of the period and thus doubling overall employment) seems to more than double end of period employment in that location. Part of the explanation why the number of initially active establishments is associated with higher terminal employment is that it increases the probability that some of these establishments become successful and grow substantially and reduces the possibility that all of them fail.[31] Estimates at higher distances are zero, except those for the 5 to 10 mile range where employment and number of firms enter with opposite signs. When including final employment and number of firms (model 6) the results are the same as in those of the conceptually similar model 4 discussed earlier.

In light of the discussion at the start of this section, the magnitude of the spatial persistence of employment appears "large." But to better assess how much larger it is compared to a benchmark of no locational advantage from any initial industry presence, we compare the results in Tables 3 and 4 with the counterpart specifications estimated at the establishment, rather than the location level. The sole modification is that in some instances we included both the location and facility initial employment levels.[32] These specifications, which are reported in Tables 5 and 6, have some similarity to employment growth regressions, except they

[31] We have verified this in our data, but do not report the results for brevity.

[32] It is important to note the the intercept of the location employment and establishment employment regressions are not comparable. If the elasticity with respect to initial employment were fixed to 1, the intercepts in the former models reflect the growth rate of software employment in a typical location, while in the latter models they reflect the growth rate of a typical establishment.

are estimated as a cross-section regression on the basis of a single employment change, rather than as a dynamic panel. An important advantage of our approach is that the elasticity estimates are directly comparable with the locational persistence of employment specifications. Another advantage of this approach is that it side-steps many of the econometric issues of estimating dynamic panel models (especially when fixed effects are included) and eliminates the need to consider entry (which is investigated separately below).

Examining model 1 of Tables 5 and 6, which is the direct counterpart of model 1 of Tables 3 and 4, we see that the estimate of the coefficient of initial establishment employment on final period establishment employment is much smaller than those obtained at the location level.[33] Increasing initial employment in an establishment by one percent leads to only 0.5 to 0.6 percent increase in final employment, whereas the corresponding figures at the location level are 0.78 to 0.84 (the higher figures in each pair are the marginal effects for the Tobit regressions). Models 2, 3 and 4 are progressively more inclusive specifications, and confirm this pattern. The results of these models corroborate the conclusion that establishment level employment persistence is lower than the location level persistence.[34]

Conceptually, the difference between firm and location employment persistence could consist of three components (i) faster growth rates of firms in locations with higher firm concentration, (ii) the capture of jobs lost to an establishment by other establishments in that same location, which can contribute to location employment growth because firm growth rates are not perfectly correlated within a location, and (iii) higher entry rates of firms in locations with prior software presence and hence the creation of more jobs by entrants in high initial employment locations. The first component is evidence of positive spillovers from co-located firms, which would yield higher end-of-period employment among firms present in a location with many other co-located firms. If that were the case, then exogenously increasing the employment level of a single establishment would lead to a smaller increase in its final employment than exogenously increasing the employment levels of all establishments in a location. Investigating this key question is possible in the context of establishment level regressions because it is possible to add the initial employment by co-located establishments in these regressions, something that is not possible in the location level regressions. We have

[33] In fact, much smaller than unity. A coefficient of 1 would imply Gibrat's law. Most of the literature investigating the premise in Gibrat's law that growth rates are independent of firm size has either found a negative association, e.g., a mean reversion effect where large firms grow slowly while small firms grow faster (see early work by Evans, 1987, and Dunne, Roberts, and Samuelson 1989, as well as later work by Hart and Oulton, 1996, and Dunne and Hughes, 1994).

[34] The number of other software establishments in the same locations seems to have no effect on final employment of an establishment, when conditioning on that establishment's initial employment. The relationship is positive and robustly significant when we do not condition on an establishment's initial employment. This confirms the findings of Holmes and Stevens (2002) who find that average establishment size is larger in localities with more establishments. In other words, an area with an industrial agglomeration measured by number of establishments in that industry is even more agglomerated when measured on the basis of employees.

re-estimated models 1-4 of Tables 5 and 6 adding as a regressor the log co-located software establishment employment (plus 1, to avoid taking the log of zero). These results, reported in columns 5-8 of Tables 5 and 6, suggest that there are no such positive synergistic effects. In fact, if anything, there seems to be a negative effect from the presence of other firms (measured by their employees) on an establishment's final period employment. When adding the number of other establishments and their employment jointly in a specification, the sign on the number of other establishments is positive (though not significant) but the coefficient on the number of employees of those establishments becomes even more strongly negative.[35] It seems, then, that the effect of initial employment on final employment does not arise from a positive effect initial employment has on the growth of the existing establishments; it arises because more of the jobs that existing establishments lose are captured by co-located establishments and because more jobs are created by entrants.

We next verify whether local capture of jobs lost by an establishment is a contributor to local employment persistence. This is done by re-estimating the regressions in Tables 3 and 4 using as the dependent variable the final employment in a location in establishments that were present at the start of the sample period (incumbent establishments). In these regressions, which are available upon request, the coefficients on initial employment are higher than those of Tables 5 and 6, indicating that the negative effect of co-location on individual establishment growth is outweighed by the tendency of jobs to remain in areas with larger prior employment.[36] Perhaps more importantly, the coefficient on initial employment when the dependent variable is the end period incumbent establishment employment is lower than the elasticities reported in Tables 3 and 4. For example, Model 1 of Table 3 yields a elasticity of 0.782, while the elasticity is 0.718 when incumbent establishment employment is the dependent variable. For the Tobit version of this model, the elasticity measured by the marginal effect is 0.844 when employment of all establishments is the dependent variable versus 0.787 when incumbent establishment employment is the dependent variable. In other words, the employment persistence in a locality is only partially driven by incumbent firms capturing jobs lost to other incumbent firms. Some of the persistence must be driven by a disproportionate entry of firms into

[35] When adding the employment of collocated software establishments, the association between initial and final establishment employment strengthens somewhat.

[36] One might at first find this paradoxical, but it not. For example, consider four establishments, A, B, C1 and C2, operating in locations A, B, and C. Suppose establishments A and C1 have initial employment of 100 while establishments B and C2 have initial employment of 50. Let there be a simple process of mean reversion, whereby the large establishments lose 10 workers who get hired by the small establishments. Let establishment C1 lose another 5 workers, so that establishment employment growth is negatively associated with location employment. Then, the elasticity of final employment with respect to initial employment is around 0.54 at the establishment level, and around 0.78 at the location level, because many jobs lost by the large establishment in the high employment location are balanced by gains in the small establishment in that same location.

locations with greater initial presence of software publishers.

In the next section, we investigate and measure the extent to which software industry activity in a location makes that location more likely to attract new software establishments.

4.2 Entry

In this section, we take a detailed look into the entry process of establishments in order to confirm its impact on the spatial persistence of industry employment. Establishment entry is defined as the introduction date of a new Enterprise Identification Number (EIN) in our dataset. Our primary aim is to understand the extent to which the prior presence of software publishing firms in a location influences the entry rate of other software firms in that same location. Since other factors are associated with entry in a particular location, care needs to be exercised in formulating an appropriate econometric design. The analysis needs to encompass a more general understanding of what factors attract a software firm to a location. As elsewhere in the paper, our analysis is reduced form, i.e., we estimate the number of entrants in a particular location as a function of location characteristics.

Most importantly, even under the null hypothesis that localization economies do not influence entry probabilities, the expected number of entrants is not uniform across all geographies defined by land area. Ideally, our unit of analysis would consist of small areas that contain suitable locations. Such suitable locations are identified when a firm from another industry that uses similar physical infrastructure as software publishing is present at that location. In principle, prior or incipient presence of such firms also indicates potential suitability of a location. In section 3.3 we define a set of non-overlapping locations that are potentially suitable for software establishment entry based on the presence of establishments in industries that have been empirically observed to share addresses with software establishments.

We estimate the expected number of software firms entering into any of these locations as a function of location characteristics similar to those used in the analysis of employment persistence. These characteristics include the number of software firms already present in that location, the number of people employed by these firms, the number of software firms (and employees) in a surrounding ring of 5 mile radius (not including the inner circle), the number of software firms and employees in a surrounding 10 mile radius ring (not including the inner 5 mile radius circle), the number of software firms and employees in a surrounding 25 mile radius ring (not including the inner 10 mile circle), the weighted number of employees by other related industry in that location (as used in the regression in section 4.1), and a number of county-level controls described in section 3.1. Not all of these characteristics are used in every regression. Of these characteristics, the number of employees by other related industry is the most important control variable for the baseline propensity of software firms to enter in that location. We emphasize that these industries are considered "related" to software firms only in terms of their similar locational requirements within an industrially fungible locality, and that they are used as proxies for the availability of facilities at a very localized scale and not in any causal sense. If software firms were picking locations within, say, a county at random, they would not be equally likely to pick every location. Locations that are appropriately zoned and more developed in terms of business infrastructure would be more likely places to locate, and this variable is used as a summary statistic of the baseline propensity by software firms to enter there. The number of software firms (and employees) in surrounding locations measures the presence of less localized spillovers.

An important difference between the employment persistence analysis and the entry analysis is that in the latter we take advantage of the time variation in our data, i.e., we estimate the number of establishments entering in a location in a particular quarter as a function of the location characteristics in the preceding quarter. Doing so does not create any econometric issues, and increases the variation we can exploit in our sample. In the typical entry case, there is a single software entrant in a given location at a given quarter, while the maximum number of entering software establishments is two (except for one occurrence of entry by 5 establishments). In particular, recalling Table 1, there are about 0.001 software entrants per location each period in the full sample, while conditional upon there having been at least one entrant in the one mile radius, the average is 1.04 new software firms in that location. Given these small counts, we have estimated the entry models using an ordered probit. Our dependent variable is the number of software publishing start-ups y in a given one mile radius location l during a given quarter t. The basic ordered probit model is

$$Y^*_{l,t} = D'_{l,t-1}\delta + R'_{l,t-1}\rho + C'_{c_l,t-1}\varphi + \tau(t) + \alpha_{c_l} + \epsilon_{lt} \tag{2}$$

where $Y^*_{l,t}$ is a continuous latent variable with two threshold points, one delineating no entry from entry by a single establishment, while the other delineating entry by one establishment from entry by two or more establishments. The independent variables can be classified into three main groups: $D_{l,t-1}$ represents software industry activity/density measures in the location, $R_{l,t-1}$ controls for other location specific variation, $C_{c_l,t-1}$ controls for county specific characteristics of the county c_l where location l is situated, and $\tau(t)$ is a quadratic function of time (the model cannot be meaningfully identified with time fixed effects). Most of the variables that may be characterized by long tails are in logs. Moreover, as elsewhere in this paper, when taking the log of the number of firms or the number of employees, the value of 1 is added. This ensures well defined regressor values when no firms are present in a location, though the coefficients are only approximately equal to elasticities (the approximation is close when the independent variable has a high mean value). In some specifications we have included county effects α_{c_l} for the top 5 counties with the most software publishers.[37] The random disturbance ϵ_{lt} has a standard normal distribution.

The results are reported in Table 7. Localized own-industry density appears to have a strong positive effect on entry probabilities, regardless of whether the number of software firms or the number of employees in those firms is used as a measure of activity. The effect appears to be stronger for distances of less than one mile, it is somewhat important for intermediate distances (1-5 miles) and is marginal or absent for distances greater than 5 miles (especially for the more comprehensive specifications). The use of county fixed effects tends to weaken the association between pre-existing software firms and subsequent entry, but only marginally, while complementing these fixed effects with time varying location characteristics tends to have no effect. Reassuringly, our control for localized activity by firms that use similar infrastructure (Other Industry Employees) is positive and strongly significant. Interestingly, the number of other software firms seems essentially equally important for the location decision of potential entrants as the total number of employees of those firms (measured by statistical significance). Moreover, the two sets of models have an approximately equally good fit as measured by the log likelihood, with the specification that uses the number of firms as the measure of software firm presence having a slight edge.

Finally, when both the number of firms and the number of their employees are used in the regression, significance drops substantially, especially for the number of employees present. For this reason, we attach greater importance to the results in the specifications (1) through (6) where either the number of establishments or the number of employees of these establishments is used as a control. From the remaining variables, high tech wages and to a lesser extent university spillovers are associated with higher entry probabilities (the former perhaps as a proxy for demand for high tech labor), and the junior college spillovers and local unemployment rates are negatively associated with entry.[38]

Despite the very small entry counts, we also estimate the entry process using the Poisson model in

[37] The remaining counties are pooled together as the excluded category. We use county effects rather than location effects to control for unobserved location heterogeneity, because multiple entries over the sample period into any particular location are rare. Moreover, the model cannot be identified with county effects for each county, without the loss of many observations.

[38] We have also re-estimated these models using ordered logit. The ordered logit parameter estimates for the agglomeration variables and their effect on entry probabilities are qualitatively similar to those of the ordered probit results. The only noticeable difference is in the rate at which spillover effects decay with distance, which tends to be smaller under the ordered logit than the ordered probit. The two sets of results are also similar with regards to the auxiliary controls, except that university spillovers are positive significant under the ordered logit.

order to investigate the sensitivity of the results to the econometric specification. This model is estimated via the pseudo Maximum Likelihood method to increase the robustness of the inference.[39] The estimated coefficients are identical to Poisson ML estimation but the standard errors are adjusted for over-dispersion, and are also clustered at the county level. Our dependent variable is the same as in the ordered probit analysis, and its conditional mean is given by

$$E[y_{lt}|\alpha_l, D_{l,t-1}, R_{l,t-1}, C_{l,t-1}] = \exp(D'_{l,t-1}\delta + R'_{l,t-1}\rho + C'_{c_l,t-1}\varphi + \tau(t) + \alpha_{c_l}) \tag{3}$$

The independent variables are classified into three main groups as in the ordered probit specification. Estimation results for these generalized Poisson regressions are contained in Table 8. The parameter estimates of the Poisson and ordered probit models are not directly comparable with each other because of differences in scaling, but statistical significance and relative magnitudes can be compared. On this basis, the Poisson and ordered probit estimates for the agglomeration variables and their effect on entry probabilities are for the most part similar. The only difference seems to be that localized own-industry density effects now appear to slightly stronger for the 1 to 5 mile distance as they are for distances of up to one mile, but only when employment levels are used as the measure of industry activity.[40]

To summarize, the presence of software establishments in an 1-mile radius location exerts a strong positive influence on the number of new software establishments entering that location, while software establishment presence in the surrounding 5 mile ring also exerts a substantial positive influence. There is a far smaller and inconsistent effect for distances further than 5 miles. These effects account for the development activity at the 1-mile resolution, and are robust to the addition of a number of other controls.

4.3 Firm Survival

In the preceding section, we have shown that locations with prior software establishment activity are more likely to attract new entrants. What has not yet been demonstrated is that these entrants last at least as long if not longer than the (fewer) entrants that choose to locate in other areas. If that were not the case, perhaps because areas with other establishments represent more fierce labor market competition, then these increased entry rates would not necessarily lead to a greater number of active establishments in the longer

[39] For a more detailed discussion of this reasoning, see Wooldridge (2002) and Cameron and Trivedi (2005). Implementation is via the PQML command in STATA with the keep option.

[40] A careful comparison between Tables 7 and 8 reveals that the specifications in the two tables are similar by not identical, thus increasing the range of models we have considered. In particular, the set of models with county fixed effects is not the same across the two tables. We have estimated using ordered probit the exact same set of specifications in Table 8, but no additional qualitative differences in the results emerge other than those already mentioned here (these results are not reported because they are largely redundant).

term. We take up the question of establishment survival in this section, where we estimate the survival rate of software establishments, allowing the hazard rate for each establishment i to vary over time as a function of the establishment's age and also be a function of time-invariant covariates. For this analysis, observations consist of all of establishments i that entered after March 2000, eliminating any concerns about left censoring, and possible selection biases that might arise from it. Right censoring is accounted for in the estimation procedure, as explained below.

Two variables are particularly important: the survival time of establishment i, χ_i (which is equal to the establishment's age at exit or the number of quarters it is observed until the end of our sample), and the exit indicator d_i that takes the value 1 when the exit date of an establishment is observed, and 0 if the observation is right censored. The set of time invariant covariates is divided into three groups, as the previous analyses, but pertains to the conditions present when the establishment enters the market, t_i. The first group, D_{l_i,t_i}, includes our usual density variables. The second group, F_{i,t_i}, includes firm specific variables such as (initial) size, and the final group C_{c_i,t_i} includes characteristics of county c_i where establishment i is located, such as the unemployment rate. In the expressions that follow, Z_{i,t_i} is used to denote all three groups and the corresponding coefficient vector is represented by ψ. In our base analysis, we adopt the Weibull specification, where the hazard rate is

$$h(\chi_i, Z_{i,t_i}) = pe^{Z_{i,t_i}\psi}\left(e^{Z_{i,t_i}\psi}\chi_i\right)^{p-1} \tag{4}$$

while the survivor function (the probability of surviving χ_i periods) is

$$S(\chi_i, Z_{i,t_i}) = e^{\left(-e^{Z_{i,t_i}\psi}\chi_i\right)^p}. \tag{5}$$

The corresponding density function of survival times is

$$\begin{aligned} f(\chi_i, Z_{i,t_i}) &= S(\chi_i, Z_{i,t_i})h(\chi_i, Z_{i,t_i}) \\ &= e^{\left(-e^{Z_{i,t_i}\psi}\chi_i\right)^p} pe^{Z_{i,t_i}\psi}\left(e^{Z_{i,t_i}\psi}\chi_i\right)^{p-1}. \end{aligned} \tag{6}$$

where $\lambda_i = e^{Z_{i,t_i}\psi}$ is the location parameter and p is the shape parameter. Having defined $h(\chi_i, Z_{i,t_i})$, $f(\chi_i, Z_{i,t_i})$, and $S(\chi_i, Z_{i,t_i})$, we now write the likelihood for the Weibull model as follows:

$$L = \prod_{i=1}^{n}\left\{f(\chi_i, Z_{i,t_i})\right\}^{d_i}\left\{S(\chi_i, Z_{i,t_i})\right\}^{1-d_i}. \tag{7}$$

where n is the number of establishments used in the duration analysis.

The estimation results are reported in Table 9. There is no association between the number of other software firms or their employment and the exit rates of an establishment, except for a very tenuous one that points to an increased hazard when an establishment enters in a location with a large number of software industry employees. But confidence in that finding is further undermined by the negative coefficient on the number of co-located establishments, which essentially cancels that effect out. In other words, the presence of other software publishers in a location does not seem to confer an advantage in terms of "fitness" (i.e., the ability to survive) via either the labor pool channel or through knowledge spillovers, and possibly confers a disadvantage. Proximate alternative employers might hasten the demise of weak firms by enabling an easy transition of their employees to other firms. Transitions between these firms can be easy not only because proximity means that these employees have small switching costs, but perhaps also because personal interactions between the employees of the two firms may facilitate the recruiting process.[41]

With regard to the other covariates, the size of the localized employment of the control industries is not expected to have any major impact on the profitability and survival of an existing establishment. Indeed, no effect is apparent for software firm exit rates. The hazard rate is increasing with establishment age ($p > 1$) but somewhat decreasing in initial establishment size.[42] However, it is important to note that firm size refers to initial size of the establishment, and an increased scale of operations may simply proxy for deeper financial resources. All other variables tend not to be significant, including the firm's initial wage. Wage rates are clearly endogenous, and they can reflect either the higher profitability of firms (more profitable firms are known to pay higher wages), or the higher quality of labor that these firms employ (which would seem to suggest that establishments with high quality labor are more likely to survive), or the need to offer higher wages to attract labor. But they seem to have no effect.

The only exception to this general non-significance is the county high-tech wage (which tends to increase exit rates) and the unemployment rate that also does so when used in conjunction with county wages. High industry wages may reflect competition for labor from other high tech industries and a high unemployment rate may reflect economic stress. To investigate the robustness of our key conclusions to the inclusion of the wage and employment levels, we have performed a series of robustness tests which are reported in Table 10. In particular, we have re-estimated many regressions with the county wage but without the own wage,

[41] An employee who is considering leaving a firm has detailed information about the new employer, while the new employer has good information about the prospective employee, reducing uncertainty and increasing the chance of a good match.

[42] For example, Dunne, Roberts and Samuelson (1988, 1989) find the failure rates of both plants and firms decreases with the size and age of the enterprise. See also related work by Dunne, Klimek and Roberts (2005) on the role of firm experience.

and have also estimated regressions without either of the two wage effects. In many of these models, the number of employees was also dropped from the model. A consistent finding is that initial establishment employment is not associated with changes in the hazard rate, but the county high tech wage is, whether or not establishment employment is included. The unemployment rate is significant only when it enters with the high-tech wage, because the two variables are negatively correlated and both enter with a positive sign.

To further evaluate the sensitivity of our findings, we also estimate a simple probit model of exit, where the vector of exogenous characteristics is defined with respect to the current period, i.e., it is time-varying. This model estimates the probability that the establishment exits following period t as a function of the variable vector $Z_{i,t}$, i.e., the same set of variables as in the Weibull duration model, except that values are not for the establishment's entry period but for the current period. The age of the establishment is also included as a regressor to allow for duration dependence. The results are qualitatively similar to the ones presented above, with only minor exceptions. The probit results do not exhibit significant duration dependence, with the log of age being positive, but not statistically significant. However, the number of employees in the preceding period reduces exit rates, probably because it reflects current profitability.

In brief, there is no strong spatial association between software publisher activity and software establishment exit rates, and there is only limited association between other variables and exit. We discuss the implications of the combination of results in the employment growth (at the location and establishment level), establishment entry, and establishment exit in our concluding section.

5 Discussion and Conclusion

Having completed and discussed each facet of the econometric analysis separately, we summarize here what the body of evidence suggests about the association between agglomeration in the software industry and the employment dynamics in that industry. In interpreting the body of findings, it is important to keep in mind that the choice of the software industry eliminates most agglomeration or localization economies other than those from human capital spillovers or through the labor market, as argued in the introduction. These sources of spatial affiliation include spin-offs, which can be considered a component of the labor market channel and are primarily relevant for entry rates. Knowledge spillovers and labor market agglomeration effects have a somewhat different "signature," as discussed at length in section 2. We will use this differential predicted effects to assess the relative importance of the two types of spillovers on the basis of our results.

At the smallest level of spatial resolution (within one mile), employment exhibits excess persistence, i.e., greater persistence than one would expect from the employment persistence of individual establishments. This persistence is not driven by the faster employment growth of establishments in areas with multiple software firms. Indeed, incumbent establishments seem to exhibit slower growth when co-located with other establishments. Moreover, the persistence is only partially driven by the fact that a job "lost" by an establishment is "captured" by another establishment in the same location. Rather, the presence of software firms increases the propensity of other software firms to enter within very close proximity. But the establishments entering in localities where other software firms are already present do not experience differential survival rates than establishments entering in localities with no prior software firm activity.

This combination of findings suggests that a prior concentration of software firms in a locality lowers the entry costs of other software firms in that same locality, but that post-entry profitability of the average firm in that locality is not higher and may even be lower (if one focuses on employment growth). Recalling the framework developed in Section 2, this pattern is most consistent with spatial effects arising from the localized labor pool, including from spin-off effects, e.g., employees leaving an existing establishment with the intention of setting-up shop in the same physical location, or firms finding it preferable to locate in a particular location because recruiting is easier. It is hard with our data to disentangle the relative importance of spin-offs but the descriptive evidence in section 3.2 argues against spin-offs (unless spin-offs enter different market segments than the parent firm). A competing (or complementary) explanation is that entrant firms co-locate because of synergies or direct human capital spillovers from the incumbent firms in those areas. However, re-examining Panel A of Figure 1 reveals that such spillovers should make the average entrant into that location more successful compared to elsewhere and should also benefit incumbent firms. There is no evidence of this effect at the 1 mile range, and therefore human capital spillovers cannot be the only source of agglomeration economies in this industry. However, some human capital spillovers might be present in conjunction with the primary labor channel effects. These human capital spillovers would reinforce the positive effects of labor channel spillovers in increasing the post-entry payoff function in high spatial agglomeration locations. Even though lower entry costs in those locations mean that software establishments co-locating with other software firms are on average weaker business prospects, the upward shift in the payoff function would leave average exit rates unchanged (and employment growth of incumbent firms only marginal lower).[43]

[43]Unfortunately, we have no way of directly estimating entry costs or firm payoffs in order to assess the extent of selection that happens due to differential entry rates and the relative contribution, if any, of human capital spillovers. However, we can certainly say that the data is inconsistent with a pure human capital spillover story and it is also inconsistent with a framework

Moving to intermediate distances (between 1 and 10 miles), we find there is no strong and consistent relationship between the end-period software employment in a location and the software employment in the surrounding area, whether initial or final (recalling the results, there is some nuance and useful insights, but no strong consistent effects). However, there is still a positive, but diminished, association between entry rates and prior software firm activity. There continues to be no association between exit rates and software firm activity. In other words, entry costs in a location appear lower if there is a labor pool in the surrounding area, but the surrounding pool does not enhance the productivity of software firms in that location. By facilitating entry, it merely leads to a reshuffling of employees from incumbent firms at the start of the sample period to the entrants. On balance, this evidence supports the labor pool interpretations we suggested in the preceding paragraph.

Finally, at even larger distances (between 10 and 25 miles), employment in a location is decreasing in the contemporaneous employment growth at those distances. This is suggestive of a "pull" effect for employees from proximate high growth areas. Also, there is now only a tenuous positive relationship between entry rates in the location and employment at 10 to 25 miles. This evidence is consistent with labor pool effects.

On net, our findings are suggestive of strong labor pool agglomeration economies, but of only weak productivity spillovers from human capital and knowledge transmission. Firms are attracted to locations because of the existing labor force, and the effect is strongest at high levels of spatial resolution. This contributes to employment growth in those locations, as does the ability of firms located in close proximity to absorb jobs lost by other co-located firms. However, firms located in those areas do not appear to grow faster, i.e., they do not benefit from the agglomeration in the industry. Indeed, it appears that the more intense competition for employees that results from this entry in areas with prior industry presence yields a more competitive labor input environment, somewhat reduces end of period employment at the firm level, and possibly slightly increases the failure rate of the entrants.

References

[1] Abramovsky, L., R. Harrison, and H. Simpson (2007). "University Research and the Location of Business R & D," *Economic Journal*, 117: C114–C141.

mentioned in footnote 5 since under that framework firm post-entry performance improves.

[2] Acs, Z. J., D. B. Audretsch, and M. P. Feldman (1994). "R&D Spillovers and Recipient Firm Size," *Review of Economics and Statistics*, 100: 336–367.

[3] Audretsch, D. B. and M. P. Feldman (1996). "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86: 630-640.

[4] Buenstorf, G. and S. Klepper (2009). "Heritage and Agglomeration: The Akron Tyre Cluster Revisited," *Economic Journal*, 119: 705-733.

[5] Ciccone, A. and R. E. Hall (1996). "Productivity and the Density of Economic Activity," *American Economic Review*, 86: 54-70.

[6] Combes, P. (2000). "Economic Structure and Local Growth: France, 1984-1993," *Journal of Urban Economics*, 47: 329-355.

[7] De Silva, D. G. and R.P. McComb (2012). "Research Universities and Regional High-Tech Firm Startups and Exit," *Economic Inquiry*, 50: 112-130.

[8] Dunne, P. and A. Hughes (1994). "Age, Size, Growth and Survival: UK Companies in the 1980s," *Journal of Industrial Economics*, 42: 115-140.

[9] Dunne, T., S. D. Klimek, and M. J. Roberts (2005). "Exit from Regional Manufacturing Markets: The Role of Entrant Experience," *International Journal of of Industrial Organization*, 23: 399-421.

[10] Dunne, T., M. J. Roberts, and L. Samuelson (1988). "Patterns of Firm Entry and Exit in the U.S. Manufacturing Industries," *Rand Journal of Economics*, 19: 495-515.

[11] Evans, D. S. (1987). "Tests of Alternative Theories of Firm Growth," *Journal of Political Economy*, 95: 657-674.

[12] Freedman, M. L. (2008). "Job Hopping, Earnings Dynamics, and industrial agglomeration in the Software Publishing Industry," *Journal of Urban Economics*, 64: 590-600.

[13] Glaeser, E. L., H. L. Kallal, J. A. Scheinkman, A. Schleifer (1992). "Growth in Cities," *Journal of Political Economy*, 100: 1126-1152.

[14] Hart, P. E. and N. Oulton (1996). "Growth and Size of Firms," *Economic Journal*, 106: 1242-1252.

[15] Henderson, J. V. (1994). "Where Does an Industry Locate?," *Journal of Urban Economics*, 35: 83-104.

[16] Henderson, V., A. Kuncoro, and M. Turner (1995). "Industrial Development in Cities," *Journal of Political Economy*, 103: 1067-1090.

[17] Holmes, T. J. and J. J. Stevens (2002). "Geographic Concentration and Establishment Scale," *Review of Economics and Statistics*, 84: 682–690.

[18] Krugman, P. (1991). "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99: 483-499.

[19] Rosenthal, S. S. and W. C. Strange (2003). "Geography, Industrial Organization, and Agglomeration," *Review of Economics and Statistics*, 85: 377-393.

[20] Saxenian, A. (1994). Regional Advantage: Culture and Competition in Silicon Valley and Route 128, Cambridge: Harvard University Press.

[21] Staber, U. (2001). "Spatial Proximity and Firm Survival in a Declining Industrial District: The Case of Knitwear Firms in Baden-Wurttemberg," *Regional Studies*, 35: 329-341.

[22] Wallsten, S. J. (2001). "An Empirical Test of Geographic Knowledge Spillovers using Geographic Information Systems and Firm-size-level Data," *Regional Science and Urban Economics*, 31: 571-599.

[23] Woodward, D., O. Figuerido, and P. Guimarães (2006). "Beyond the SiliconValley: Univeristy R&D and high technology location," *Journal of Urban Economics*, 60: 15-32.

Table 1: Summary statistics for software firms

Panel A: Full sample of firms			
	All	Incumbents	Entrants[a]
Unique number of firms	877	506	371
Number of establishments	1.228	1.106	1.552
per firm	(1.041)	(0.583)	(1.707)
Unique number of establishments	957	526	431
Number of firms with	757	497	360
one establishment			
Average size (employment)	34.955	31.071	45.256
of establishments	(194.517)	(113.474)	(322.322)
Own quarterly wage ($) per	20,305.82	19,089.65	23,530.49
establishment	(25,229.45)	(20,133.96)	(35,147.05)
Age (in months) per establishment	105.983	129.217	44.379
	(85.166)	(83.681)	(51.886)
Average time in the sample (quarters)	11.7000	13.038	8.150
	(7.6111)	(7.758)	(5.880)
University spillover ($)	3,379,909.00	3,422,106.00	3,268,024.00
	(9,146,528.00)	(9,487,206.00)	(8,174,685.00)
Junior college spillover ($)	16,301.18	18,623.02	10,144.80
	(65,706.70)	(75,875.25)	(21,171.85)
Quarterly average wage ($) rate for	13,653.58	13,433.91	14,236.03
high-tech industries in the county	(3,717.31)	(3,676.62)	(3,761.90)
County unemployment rate	5.385	5.362	5.445
	(1.239)	(1.284)	(1.109)
County amenity LQ	1.065	1.066	1.064
	(0.198)	(0.202)	(0.188)
Panel B: Randomly chosen non-overlapping locations			
	All	At least one incumbent	At least one entrant
Number of unique locations	9,299	201	170
Average number of software	1.478	65.870	68.064
employees: $1 \leq$ mile	(34.464)	(224.785)	(241.105)
Average number of software	33.078	661.003	859.977
employees: > 1 - $5 \leq$ miles	(284.603)	(1,241.156)	(1,400.822)
Average number of other industry	646.724	10,718.010	11,798.320
employees: $1 \leq$ mile	(3,828.118)	(19,788.580)	(21,665.510)
Average number of other industry	11,985.390	138,240.400	144,071.500
employees: > 1 - $5 \leq$ miles	(42,743.660)	(141,583.400)	(145,221.700)
Average number of software	0.001	0.033	0.063
entrants	(0.036)	(0.195)	(0.259)
Average size of other establishments	11.281	24.882	28.495
	(32.011)	(37.7587)	(43.378)
Average quarterly wage ($) other	6,058.15	11,028.08	12,035.130
establishments	(10,669.55)	(5,369.81)	(5,676.845)

Standard deviations are in parentheses.

[a] Entered after Mar 31, 2000

Table 2: Rival information within five miles.

City	Location	Establishment	Category	Entry date	Exit date	Quarterly employment	Distance (in miles)
A	1	1	Comp. software & services.	1990-03	2005-02	91.14	–
		2	Interactive comp. software.	2005-02		141.79	–
	2	3	Comp., peripheral equip.and software.	2000-11		2.44	1.005
	3	4	Business software.	1997-01		39.90	2.030
	4	5	Aircraft performance software.	1996-10		3.01	3.746
	5	6	Develop comp. software.	1977-12		33.56	3.834
B	1	1	Electronic trading software.	1993-12	2005-09	22.722	
	2	2	Investment & banking software.	1983-12		5.890	0.579
		3	Comp. software & services.	1992-04		68.32	0.579
	3	4	Accounting services software.	1996-04	2005-12	3.067	0.587
	4	5	Geophysical and Meteorological software.	1991-03	2005-12	5.967	0.732
	5	6	Healthcare software.	1990-12		46.667	0.943
	6	7	Software engineering.	1993-06		5.528	1.004
C	1	1	Comp. & software related services.	2000-12		3.81	–
	2	2	Comp. software.	1998-04	2005-12	14.58	0.69
		3	Comp. software services.	2006-01		13.08	0.69
	3	4	Comp. consultants.	1999-11		7.88	1.79
	4	5	No information.	1990-09	2004-03	2.00	3.39
	5	6	Industrial control instruments software.	1990-12		3.90	3.41
	6	7	Coommunication systems software.	1992-09		340.85	3.53
	7	8	Systems software development services.	2004-02		16.33	3.84
	8	9	Data management and business software.	1994-03		27.38	3.93
	9	10	Fluid meters & counting devices software.	2001-11		12.31	4.30
	10	11	Customer privacy & protection.	1995-01		652.10	4.51
	11	12	Comp. software services.	2003-01		148.50	4.58
	12	13	Coommunication systems software.	1997-03		5.77	4.64
	13	14	Identity management solutions.	1990-09		5.56	4.75
		15	Engineering & communication software.	1996-01		32.55	4.75
	14	16	Business & engineering software.	1998-10	2004-03	3.08	4.89
D	1	1	Business solutions.	1999-01		7.264	
	2	2	Insurance and business software.	1999-01	2006-06	572.00	0.56
		3	Business intelligence software.	2002-10		598.500	0.56
	3	4	Information technology.	1996-02		2.00	3.17
	4	5	No information	2002-01		7.417	4.86
E	1	1	Entertainment software.	1998–07		18.15	–
		2	Develop comp. games software.	2004-01	2005-09	9.14	–
	2	3	Software for automating data integration.	1991-03		54.75	0.12
	3	4	Entertainment software.	2002-09	2002-12	4.00	0.17
	4	5	Engineering software.	2001-02	2005-03	6.10	0.25
		6	Enterprise network configurations.	2001-10		48.28	0.25
	5	7	Advertising software.	1999-04	2000-12	42.69	0.39
	6	8	Creative design software.	2000-12	2006-06	6.35	0.49
	7	9	Digital solutions software.	1994-07	2006-09	2.28	0.49
	8	10	Programming, analysing, & designing.	1990-09		2.30	0.62

Table 3: Employment growth at a random location

Variable	Log(*employment* + 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log initial number of software	0.782***	0.774***	0.664***	0.696***	0.630***	0.641***
employees: 1 $\leq$ mile	(0.080)	(0.081)	(0.091)	(0.088)	(0.111)	(0.108)
Log initial number of software		0.072	0.005	0.166*	0.033	0.206
employees: $>$ 1 - 5 $\leq$ miles		(0.049)	(0.051)	(0.087)	(0.118)	(0.138)
Log initial number of software		-0.141	-0.165**	-0.122	-0.513**	-0.362*
employees: $>$ 5 - 10 $\leq$ miles		(0.086)	(0.082)	(0.148)	(0.207)	(0.219)
Log initial number of software		0.115	0.101	0.418**	-0.061	0.210
employees: $>$ 10 - 25 $\leq$ miles		(0.076)	(0.075)	(0.168)	(0.135)	(0.205)
Log initial number of other industry			0.107	0.138	0.111	0.170
employees: 1 $\leq$ mile			(0.168)	(0.189)	(0.185)	(0.200)
Log initial number of software					0.606	0.763*
firms: 1 $\leq$ mile					(0.406)	(0.397)
Log initial number of software					-0.164	0.117
firms: $>$ 1 - 5 $\leq$ miles					(0.307)	(0.513)
Log initial number of software					0.728*	0.992**
firms: $>$ 5 - 10 $\leq$ miles					(0.393)	(0.501)
Log initial number of software					0.293	0.848
firms: $>$ 10 - 25 $\leq$ miles					(0.333)	(0.583)
Log number of rival software						-0.080
firms: $>$ 1 - 5 $\leq$ miles$_{T-1}$						(0.454)
Log number of rival software						-0.632
firms: $>$ 5 - 10 $\leq$ miles$_{T-1}$						(0.611)
Log number of rival software						-0.183
firms: $>$ 10 – 25 $\leq$ miles$_{T-1}$						(0.633)
Rivals' log of number of software				-0.211**		-0.288*
employees: $>$ 1 – 5 $\leq$ miles$_{T-1}$				(0.086)		(0.148)
Rivals' log of number of software				-0.034		-0.030
employees: $>$ 5 – 10 $\leq$ miles$_{T-1}$				(0.132)		(0.197)
Rivals' log of number of software				-0.382**		-0.435*
employees: $>$ 10 – 25 $\leq$ miles$_{T-1}$				(0.171)		(0.232)
Log number of other industry			0.132	0.134	0.073	0.045
employees: 1 $\leq$ mile$_{T-1}$			(0.180)	(0.201)	(0.205)	(0.211)
Initial variables		Yes**	Yes**	Yes**	Yes**	Yes**
Lagged variables $_{T-1}$				Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Number of obs.	201	201	201	201	201	201
Adj R^2	0.499	0.520	0.549	0.571	0.571	0.599

*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level. Robust standard errors are in parentheses. The dependent variable is the log of total number of employees in non-overlapping rings in the last period. Initial and lagged variables include spillovers, unemployment rate, and county high-tech wages.

Table 4: Employment growth at a random location: Tobit regression results

Variable	Log(*employment* + 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log initial number of software	0.861***	0.857***	0.732***	0.769***	0.701***	0.714***
employees: $1 \leq$ mile	(0.083)	(0.088)	(0.100)	(0.095)	(0.121)	(0.115)
Log initial number of software		0.074	-0.007	0.211*	-0.004	0.196
employees: > 1 - $5 \leq$ miles		(0.067)	(0.070)	(0.121)	(0.159)	(0.179)
Log initial number of software		-0.187	-0.214*	-0.138	-0.732**	-0.482
employees: > 5 - $10 \leq$ miles		(0.116)	(0.111)	(0.216)	(0.285)	(0.303)
Log initial number of software		0.156	0.139	0.537**	-0.037	0.266
employees: > 10 - $25 \leq$ miles		(0.101)	(0.101)	(0.211)	(0.167)	(0.243)
Log initial number of other industry			0.166	0.196	0.169	0.233
employees: $1 \leq$ mile			(0.226)	(0.245)	(0.238)	(0.252)
Log initial number of software					0.621	0.796*
firms: $1 \leq$ mile					(0.451)	(0.438)
Log initial number of software					-0.140	0.375
firms: > 1 - $5 \leq$ miles					(0.392)	(0.661)
Log initial number of software					1.071**	1.493**
firms: > 5 - $10 \leq$ miles					(0.532)	(0.712)
Log initial number of software					0.265	1.349*
firms: > 10 - $25 \leq$ miles					(0.408)	(0.738)
Log number of rival software						-0.287
firms: $> 1 - 5 \leq \text{miles}_{T-1}$						(0.575)
Log number of rival software						-0.796
firms: $> 5 - 10 \leq \text{miles}_{T-1}$						(0.812)
Log number of rival software						-0.614
firms: $> 10 - 25 \leq \text{miles}_{T-1}$						(0.781)
Rivals' log of number of software				-0.276**		-0.332*
employees: $> 1 - 5 \leq \text{miles}_{T-1}$				(0.116)		(0.183)
Rivals' log of number of software				-0.091		-0.147
employees: $> 5 - 10 \leq \text{miles}_{T-1}$				(0.197)		(0.255)
Rivals' log of number of software				-0.493**		-0.504*
employees: $> 10 - 25 \leq \text{miles}_{T-1}$				(0.219)		(0.283)
Log number of other industry			0.111	0.130	0.050	0.026
employees: $1 \leq \text{mile}_{T-1}$			(0.249)	(0.267)	(0.270)	(0.275)
Initial variables		Yes**	Yes**	Yes**	Yes**	Yes**
Lagged variables $_{T-1}$				Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Marginal effect of log initial number of software employees: $1 \leq$ mile	0.844	0.839	0.717	0.754	0.687	0.698
Number of obs.	201	201	201	201	201	201
Number of uncensored obs.	150	150	150	150	150	150
Log likelihood	-340.748	-335.060	-330.038	-324.306	-324.615	-318.103

*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level.

The dependent variable is the log of total number of employees in non-overlapping rings in the last period.

Initial and lagged variables include spillovers, unemployment rate, and county high-tech wages.

Table 5: Firm-level employment growth for incumbents

Variable	Establishment-level log($employment$ + 1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of initial establishment-level	0.498***	0.493***	0.478***	0.482***	0.524***	0.525***	0.517***	0.531***
employment	(0.069)	(0.070)	(0.073)	(0.073)	(0.072)	(0.073)	(0.074)	(0.075)
Log initial number of other software					-0.046	-0.060*	-0.102***	-0.144*
employees: 1 $\leq$ mile					(0.031)	(0.035)	(0.039)	(0.077)
Log initial number of other software		-0.037	-0.055	0.012		-0.013	-0.031	-0.039
employees: $>$ 1 - 5 $\leq$ miles		(0.036)	(0.038)	(0.090)		(0.038)	(0.039)	(0.096)
Log initial number of other software		0.084	0.088	0.126		0.076	0.078	0.091
employees: $>$ 5 - 10 $\leq$ miles		(0.054)	(0.055)	(0.129)		(0.055)	(0.054)	(0.130)
Log initial number of other software		-0.068	-0.061	-0.031		-0.075	-0.065	-0.061
employees: $>$ 10 - 25 $\leq$ miles		(0.052)	(0.052)	(0.156)		(0.052)	(0.051)	(0.156)
Log initial number of other industry			0.045	0.066*			0.088**	0.086**
employees: 1 $\leq$ mile			(0.037)	(0.039)			(0.041)	(0.041)
Log initial number of other software				-0.103				0.120
firms: 1 $\leq$ mile				(0.109)				(0.183)
Log initial number of other software				-0.103				-0.002
firms: $>$ 1 - 5 $\leq$ miles				(0.190)				(0.196)
Log initial number of other software				-0.070				-0.030
firms: $>$ 5 - 10 $\leq$ miles				(0.256)				(0.256)
Log initial number of other software				-0.054				0.005
firms: $>$ 10 - 25 $\leq$ miles				(0.259)				(0.260)
Initial variables		Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Number of obs.	528	528	528	528	528	528	528	528
Adj R^2	0.175	0.182	0.184	0.187	0.177	0.185	0.192	0.193

*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level. Robust standard errors are in parentheses. The dependent variable is the log of total number of employees in an establishment in the last period. Initial variables include spillovers, unemployment rate, and county high-tech wages.

Table 6: Firm-level employment growth for incumbents: Tobit regression results

Variable	Establishment-level log($employment$ + 1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of initial establishment-level	0.700***	0.694***	0.656***	0.663***	0.774***	0.766***	0.749***	0.781***
employment	(0.092)	(0.092)	(0.095)	(0.095)	(0.101)	(0.102)	(0.102)	(0.108)
Log initial number of other software					-0.123*	-0.135*	-0.232***	-0.332**
employees: 1 $\leq$ mile					(0.070)	(0.080)	(0.089)	(0.141)
Log initial number of other software		-0.103	-0.153*	-0.038		-0.048	-0.094	-0.152
employees: $>$ 1 - 5 $\leq$ miles		(0.074)	(0.080)	(0.195)		(0.081)	(0.082)	(0.200)
Log initial number of other software		0.167	0.174	0.242		0.149	0.149	0.165
employees: $>$ 5 - 10 $\leq$ miles		(0.109)	(0.109)	(0.259)		(0.109)	(0.109)	(0.260)
Log initial number of other software		-0.110	-0.087	-0.065		-0.123	-0.092	-0.128
employees: $>$ 10 - 25 $\leq$ miles		(0.101)	(0.102)	(0.334)		(0.101)	(0.102)	(0.334)
Log initial number of other industry			0.125	0.168**			0.216***	0.212**
employees: 1 $\leq$ mile			(0.076)	(0.082)			(0.083)	(0.084)
Log initial number of other software				-0.256				0.251
firms: 1 $\leq$ mile				(0.208)				(0.298)
Log initial number of other software				-0.128				0.099
firms: $>$ 1 - 5 $\leq$ miles				(0.404)				(0.414)
Log initial number of other software				-0.164				-0.069
firms: $>$ 5 - 10 $\leq$ miles				(0.514)				(0.513)
Log initial number of other software				-0.028				0.089
firms: $>$ 10 - 25 $\leq$ miles				(0.546)				(0.545)
Initial variables		Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Marginal effect of log of initial establishment-level employment	0.574	0.569	0.535	0.541	0.636	0.628	0.611	0.636
Number of obs.	528	528	528	528	528	528	528	528
Number of uncensored obs.	268	268	268	268	268	268	268	268
Log likelihood	-807.024	-804.081	-802.723	-801.607	-805.474	-802.666	-799.277	-798.824

*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level. Robust standard errors are in parentheses. The dependent variable is the log of total number of employees in an establishment in the last period. Initial variables include spillovers, unemployment rate, and county high-tech wages.

Table 7: Ordered probit regression results for software entry

Variable	Number of new software entrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of number of software	0.372***	0.328***	0.352***				0.264***
firms: 1 $\leq$mile$_{t-1}$	(0.047)	(0.049)	(0.048)				(0.088)
Log of number of software	0.205***	0.192***	0.178***				0.015
firms: $>$ 1 – 5 $\leq$ miles$_{t-1}$	(0.042)	(0.043)	(0.042)				(0.071)
Log of number of software	0.041	0.040	0.020				0.217**
firms: $>$ 5 – 10 $\leq$ miles$_{t-1}$	(0.048)	(0.049)	(0.048)				(0.097)
Log of number of software	0.083**	0.095***	0.010				-0.105
firms: $>$ 10 – 25 $\leq$ miles$_{t-1}$	(0.034)	(0.035)	(0.035)				(0.072)
Log of number of software				0.119***	0.104***	0.111***	0.035
employees: 1 $\leq$ mile$_{t-1}$				(0.015)	(0.015)	(0.015)	(0.029)
Log of number of software				0.111***	0.100***	0.100***	0.077***
employees: $>$ 1 – 5 $\leq$ miles$_{t-1}$				(0.016)	(0.017)	(0.016)	(0.029)
Log of number of software				0.001	0.002	-0.016	-0.099**
employees: $>$ 5 – 10 $\leq$ miles$_{t-1}$				(0.019)	(0.019)	(0.019)	(0.044)
Log of number of software				0.058***	0.060***	0.025	0.060*
employees: $>$ 10 – 25 $\leq$ miles$_{t-1}$				(0.015)	(0.015)	(0.016)	(0.035)
Log of other industry	0.119***	0.143***	0.127***	0.127***	0.152***	0.134***	0.123***
employees: 1 $\leq$ mile$_{t-1}$	(0.017)	(0.018)	(0.017)	(0.016)	(0.017)	(0.017)	(0.018)
Log of university spillover$_{t-1}$			0.021			0.023*	0.019
			(0.013)			(0.013)	(0.013)
Log of junior college spillover$_{t-1}$			-0.027*			-0.027*	-0.024
			(0.016)			(0.016)	(0.016)
Log of average wage of high-tech			0.753***			0.774***	0.763***
industries in the county$_{t-1}$			(0.127)			(0.124)	(0.128)
County unemployment rate$_{t-1}$			-0.078***			-0.073***	-0.076***
			(0.028)			(0.028)	(0.028)
County amenity LQ_{t-1}			-0.009			-0.051	-0.025
			(0.107)			(0.115)	(0.110)
Trend	-0.359	-0.382	0.458	-0.167	-0.193	0.519	0.463
	(0.339)	(0.341)	(0.491)	(0.339)	(0.341)	(0.493)	(0.495)
Trend^2	-0.049	-0.045	-0.928**	-0.252	-0.241	-1.003**	-0.946**
	(0.332)	(0.334)	(0.464)	(0.331)	(0.333)	(0.466)	(0.468)
Top 5 county effects		Yes**			Yes**		
Thresholds							
μ_1	3.651***	3.683***	10.045***	3.733***	3.749***	10.300***	10.148***
	(0.084)	(0.086)	(1.139)	(0.086)	(0.088)	(1.120)	(1.150)
μ_2	4.954***	5.002***	11.382***	5.030***	5.061***	11.630***	11.490***
	(0.138)	(0.140)	(1.149)	(0.139)	(0.141)	(1.129)	(1.159)
Number of obs.	251,073	251,073	251,073	251,073	251,073	251,073	251,073
Log likelihood	-1,408.115	-1,394.585	-1,383.135	-1,412.280	-1,401.223	-1,386.207	-1376.097
LR χ^2	1,616.660	1,643.710	1,666.620	1,608.330	1,630.4440	1,660.470	1680.690
Pseudo R^2	0.365	0.371	0.376	0.363	0.368	0.375	0.379

*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level. Robust standard errors are in parentheses. The dependent variable takes the value 0 for no entrants, 1 for one entrant, and 2 for two or more entrants at a given location.

Table 8: Poisson regression results for software entry

Variable	Number of new software entrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of number of software	0.622***	0.625***	0.663***				0.468**
firms: $1 \leq \text{mile}_{t-1}$	(0.119)	(0.118)	(0.115)				(0.210)
Log of number of software	0.545***	0.501***	0.460***				-0.003
firms: $> 1 - 5 \leq \text{miles}_{t-1}$	(0.122)	(0.120)	(0.124)				(0.196)
Log of number of software	0.190	0.167	0.094				0.504**
firms: $> 5 - 10 \leq \text{miles}_{t-1}$	(0.151)	(0.146)	(0.148)				(0.253)
Log of number of software	0.352***	0.166	0.092				-0.158
firms: $> 10 - 25 \leq \text{miles}_{t-1}$	(0.107)	(0.107)	(0.110)				(0.197)
Log of number of software				0.204***	0.204***	0.212***	0.082
employees: $1 \leq \text{mile}_{t-1}$				(0.034)	(0.033)	(0.031)	(0.064)
Log of number of software				0.285***	0.269***	0.270***	0.227***
employees: $> 1 - 5 \leq \text{miles}_{t-1}$				(0.046)	(0.047)	(0.049)	(0.083)
Log of number of software				0.062	0.035	-0.006	-0.203*
employees: $> 5 - 10 \leq \text{miles}_{t-1}$				(0.053)	(0.053)	(0.054)	(0.115)
Log of number of software				0.174***	0.096**	0.073	0.123
employees: $> 10 - 25 \leq \text{miles}_{t-1}$				(0.046)	(0.045)	(0.047)	(0.086)
Log of other industry	0.371***	0.363***	0.331***	0.393***	0.385***	0.348***	0.316***
employees: $1 \leq \text{mile}_{t-1}$	(0.049)	(0.049)	(0.049)	(0.044)	(0.045)	(0.044)	(0.051)
Log of university spillover$_{t-1}$		0.011	0.087***		0.019	0.075**	0.072**
		(0.043)	(0.033)		(0.041)	(0.033)	(0.033)
Log of junior college spillover$_{t-1}$		-0.088*	-0.095**		-0.074	-0.084*	-0.078
		(0.052)	(0.048)		(0.052)	(0.048)	(0.049)
Log of average wage of high-tech		3.165***	2.367***		3.230***	2.452***	2.421***
industries in the county$_{t-1}$		(0.290)	(0.331)		(0.288)	(0.327)	(0.328)
County unemployment rate$_{t-1}$		0.056	-0.206***		0.051	-0.184**	-0.193**
		(0.075)	(0.074)		(0.074)	(0.074)	(0.075)
County amenity LQ_{t-1}		0.174	-0.114		0.119	-0.251	-0.169
		(0.290)	(0.289)		(0.296)	(0.317)	(0.304)
Trend	-1.227	-2.237	1.628	-0.516	-1.648	1.792	1.568
	(0.839)	(1.391)	(1.417)	(0.839)	(1.375)	(1.418)	(1.418)
Trend^2	0.125	0.319	-2.799**	-0.571	-0.208	-2.999**	-2.803**
	(0.800)	(1.262)	(1.286)	(0.797)	(1.247)	(1.286)	(1.284)
Top 5 county effects	Yes**	Yes**		Yes**	Yes**		
Number of obs.	251,073	251,073	251,073	251,073	251,073	251,073	251,073
Log likelihood	-1,416.075	-1,377.211	-1,396.802	-1,423.232	-1,382.315	-1,397.364	-1,390.025

*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level. Robust standard errors are in parentheses. The dependent variable is the number of new software entrants.

All models are estimated using STATA's ppml routine with keep option.

Table 9: Survival estimates

Variable	Hazard rate determinants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of number of rival software	0.013	-0.006			-0.264	-0.267	-0.212
firms: $1 \leq \text{mile}_{t=1}$	(0.137)	(0.137)			(0.210)	(0.209)	(0.215)
Log of number of rival software	-0.089	-0.086			-0.036	-0.038	-0.119
firms: $> 1 - 5 \leq \text{mile}_{t=1}$	(0.137)	(0.136)			(0.263)	(0.263)	(0.270)
Log of number of rival software	0.170	0.171			0.050	0.047	0.017
firms: $> 5 - 10 \leq \text{mile}_{t=1}$	(0.151)	(0.151)			(0.289)	(0.288)	(0.294)
Log of number of rival software	-0.018	-0.021			-0.156	-0.158	-0.263
firms: $> 10 - 25 \leq \text{mile}_{t=1}$	(0.101)	(0.103)			(0.285)	(0.286)	(0.294)
Rivals' log of number of software			0.032	0.053	0.127*	0.127*	0.105
employees: $1 \leq \text{mile}_{t=1}$			(0.048)	(0.048)	(0.072)	(0.072)	(0.075)
Rivals' log of number of software			-0.045	-0.045	-0.030	-0.029	0.001
employees: $> 1 - 5 \leq \text{mile}_{t=1}$			(0.058)	(0.058)	(0.118)	(0.118)	(0.121)
Rivals' log of number of software			0.066	0.061	0.055	0.056	0.067
employees: $> 5 - 10 \leq \text{mile}_{t=1}$			(0.065)	(0.065)	(0.129)	(0.128)	(0.129)
Rivals' log of number of software			0.011	0.015	0.098	0.099	0.129
employees: $> 10 - 25 \leq \text{mile}_{t=1}$			(0.063)	(0.063)	(0.172)	(0.172)	(0.175)
Log of other industry	-0.030	-0.004	-0.044	-0.031	-0.006	-0.005	-0.007
employees: $1 \leq \text{mile}_{t=1}$	(0.060)	(0.060)	(0.057)	(0.056)	(0.060)	(0.060)	(0.060)
Log of university spillover$_{t=1}$	0.009	0.006	0.012	0.009	-0.008	-0.007	-0.014
	(0.061)	(0.061)	(0.062)	(0.062)	(0.062)	(0.062)	(0.055)
Log of junior college spillover$_{t=1}$	0.018	0.015	0.015	0.011	0.011	0.012	0.022
	(0.046)	(0.046)	(0.047)	(0.047)	(0.047)	(0.046)	(0.045)
Log number of employees$_{t=1}$		-0.101		-0.113	-0.150*	-0.151*	-0.132*
		(0.074)		(0.075)	(0.079)	(0.080)	(0.080)
Log of quarterly own wage$_{t=1}$	-0.033	-0.015	-0.041	-0.029	-0.017		
	(0.093)	(0.090)	(0.093)	(0.090)	(0.091)		
Log of average wage of high-tech							1.585***
industries in the county$_{t=1}$							(0.581)
County unemployment rate$_{t=1}$	0.123	0.130	0.129	0.136	0.148	0.150	0.264**
	(0.119)	(0.119)	(0.118)	(0.118)	(0.119)	(0.119)	(0.124)
County amenity $LQ_{t=1}$	-1.208**	-1.203**	-1.280**	-1.321**	-1.316**	-1.304***	-1.473***
	(0.513)	(0.504)	(0.524)	(0.518)	(0.517)	(0.506)	(0.514)
Initial trend & initial trend^2	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Top 5 county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Hazard rate: $\ln(p)$	0.361***	0.366***	0.361***	0.367***	0.369***	0.369***	0.387***
	(0.063)	(0.062)	(0.063)	(0.064)	(0.063)	(0.063)	(0.064)
Number of obs.	423	423	423	423	423	423	423
Log likelihood	-395.528	-394.228	-395.191	-393.623	-392.456	-392.474	-388.771
Wald χ^2	35.790	37.290	36.260	38.150	40.780	40.750	53.750

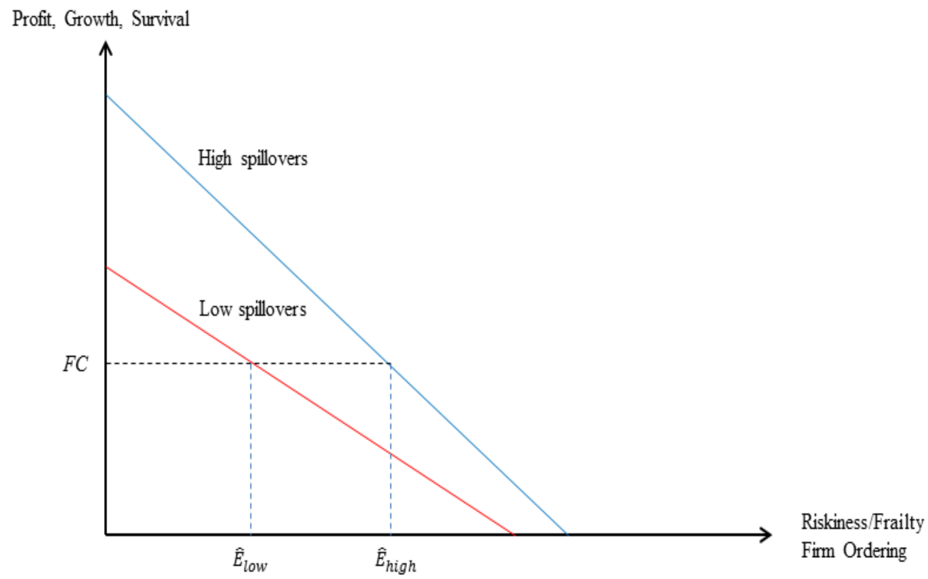
*** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percent level and * denotes statistical significance at the 10 percent level. Robust standard errors are in parentheses.

Table 10: Sensitivity analysis for survival estimates

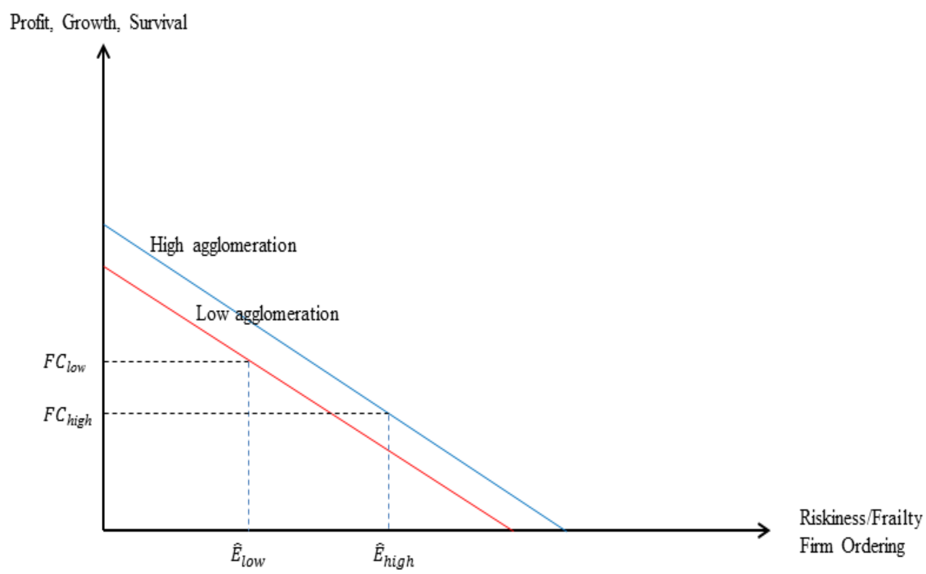
Variable	Hazard rate determinants							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of number of rival software	0.007	-0.009			0.008	-0.006		
firms: $1 \leq \text{mile}_{t=1}$	(0.136)	(0.136)			(0.137)	(0.137)		
Log of number of rival software	-0.089	-0.086			-0.115	-0.110		
firms: $> 1 - 5 \leq \text{miles}_{t=1}$	(0.137)	(0.136)			(0.133)	(0.132)		
Log of number of rival software	0.167	0.169			0.157	0.157		
firms: $> 5 - 10 \leq \text{miles}_{t=1}$	(0.151)	(0.151)			(0.151)	(0.150)		
Log of number of rival software	-0.019	-0.022			-0.084	-0.079		
firms: $> 10 - 25 \leq \text{miles}_{t=1}$	(0.102)	(0.103)			(0.104)	(0.105)		
Rivals' log of number of software			0.029	0.051			0.020	0.039
employees: $1 \leq \text{mile}_{t=1}$			(0.048)	(0.048)			(0.049)	(0.048)
Rivals' log of number of software			-0.045	-0.046			-0.046	-0.046
employees: $> 1 - 10 \leq \text{miles}_{t=1}$			(0.058)	(0.058)			(0.056)	(0.056)
Rivals' log of number of software			0.065	0.059			0.051	0.047
employees: $> 1 - 10 \leq \text{miles}_{t=1}$			(0.065)	(0.065)			(0.064)	(0.065)
Rivals' log of number of software			0.010	0.014			-0.023	-0.015
employees: $> 10 - 25 \leq \text{miles}_{t=1}$			(0.062)	(0.063)			(0.063)	(0.064)
Log of other industry	0.008	0.027	-0.001	0.007	0.009	0.025	0.001	0.009
employees: $1 \leq \text{mile}_{t=1}$	(0.057)	(0.057)	(0.054)	(0.053)	(0.056)	(0.056)	(0.053)	(0.052)
Log of university $\text{spillover}_{t=1}$	0.010	0.007	0.013	0.010	0.002	-0.001	0.005	0.002
	(0.061)	(0.061)	(0.062)	(0.062)	(0.054)	(0.053)	(0.055)	(0.055)
Log of junior college $\text{spillover}_{t=1}$	0.021	0.017	0.019	0.014	0.031	0.027	0.026	0.022
	(0.045)	(0.045)	(0.046)	(0.046)	(0.044)	(0.045)	(0.045)	(0.045)
Log number of $\text{employees}_{t=1}$		-0.102		-0.114		-0.089		-0.098
		(0.075)		(0.076)		(0.075)		(0.076)
Log of average wage of high-tech					1.722***	1.651***	1.645***	1.556***
industries in the $\text{county}_{t=1}$					(0.567)	(0.565)	(0.564)	(0.565)
County unemployment $\text{rate}_{t=1}$	0.126	0.131	0.133	0.139	0.247**	0.250**	0.254**	0.257**
	(0.118)	(0.118)	(0.117)	(0.117)	(0.123)	(0.123)	(0.124)	(0.124)
County amenity $LQ_{t=1}$	-1.185**	-1.193**	-1.249**	-1.300**	-1.393***	-1.393***	-1.468***	-1.508***
	(0.501)	(0.493)	(0.513)	(0.507)	(0.511)	(0.503)	(0.519)	(0.513)
Initial trend & initial trend^2	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Top 5 county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Hazard rate: $\ln(p)$	0.361***	0.366***	0.361***	0.367***	0.382***	0.385***	0.381***	0.384***
	(0.062)	(0.063)	(0.062)	(0.063)	(0.063)	(0.064)	(0.063)	(0.064)
Number of obs.	423	423	423	423	423	423	423	423
Log likelihood	-395.594	-394.242	-395.292	-393.673	-391.093	-390.073	-391.121	-389.943
Wald χ^2	35.630	37.230	36.060	38.090	48.670	50.510	48.220	49.910

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

Figure 1: Profitability and entry equilibrium in high and low spillover locations

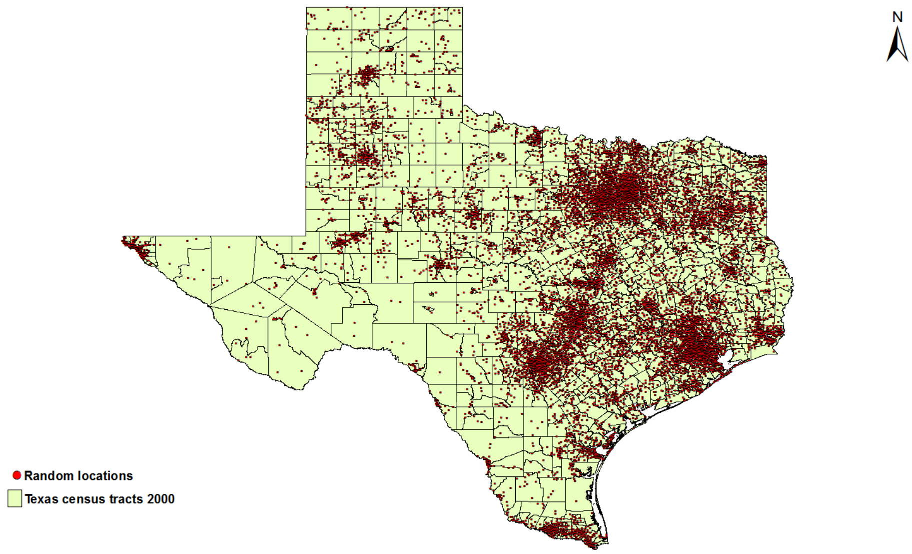


Panel A: High and low knowledge spillover locations

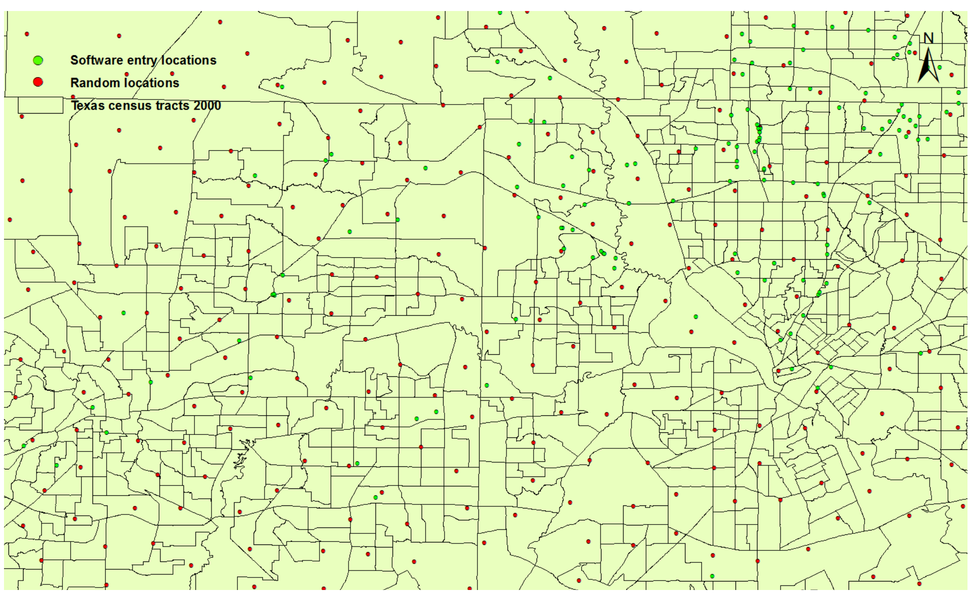


Panel B: High and low labor market agglomeration economies

Figure 2: Non-overlapping locations

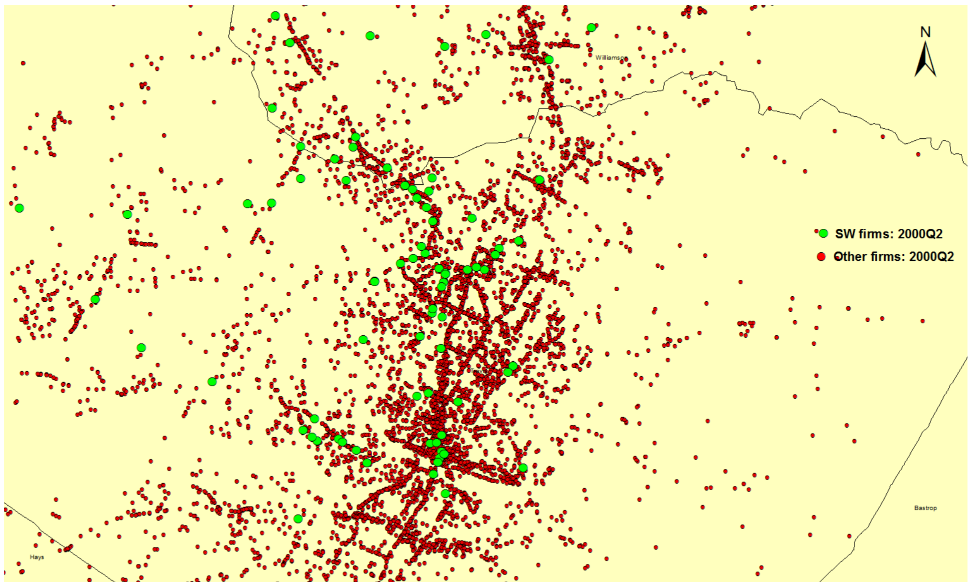


Panel A: Texas

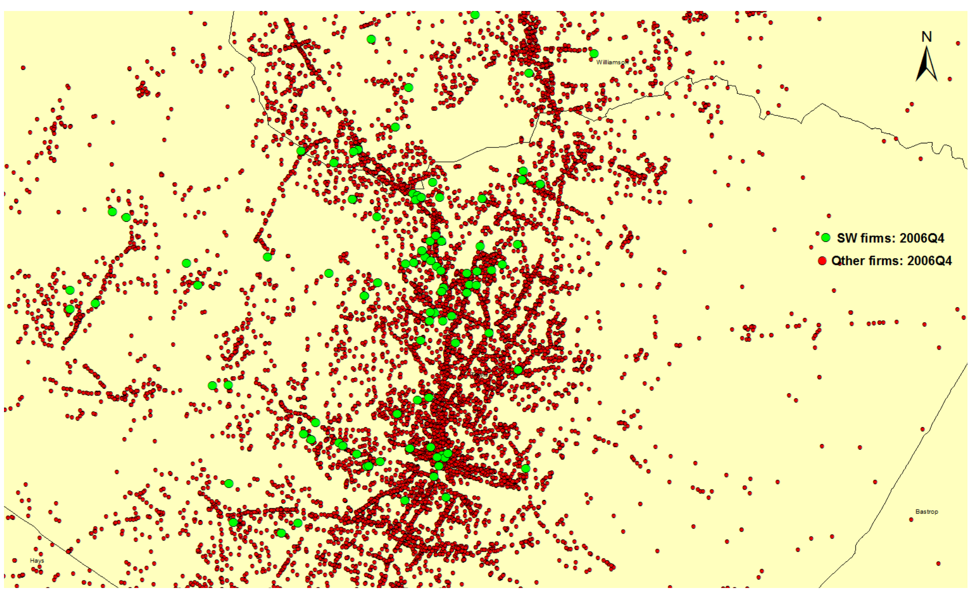


Panel B: Dallas – Fort Worth

Figure 3: Distribution of Software and Other Firms in Austin

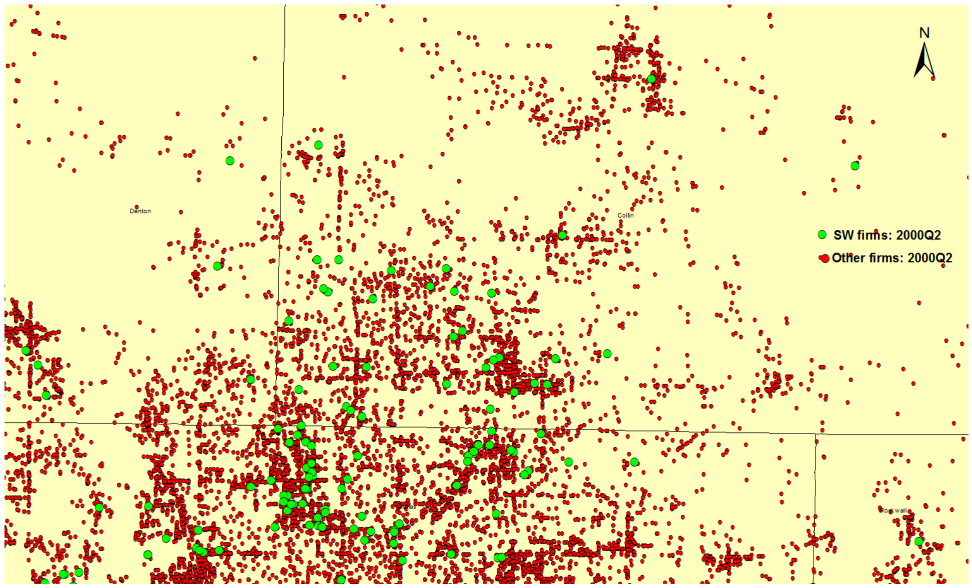


Panel A: 2000 Q2

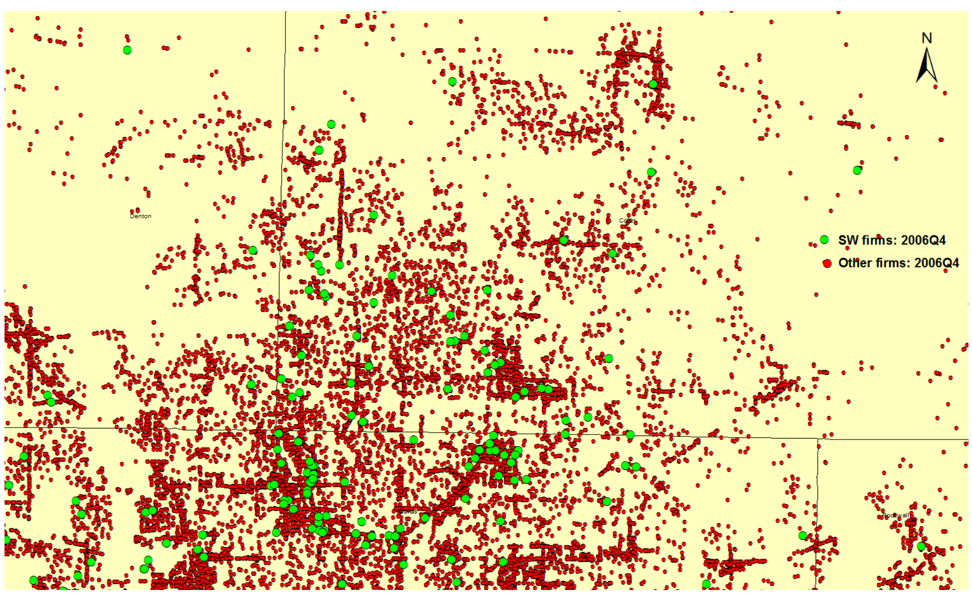


Panel B: 2006 Q4

Figure 4: Distribution of Software and Other Firms in North Dallas

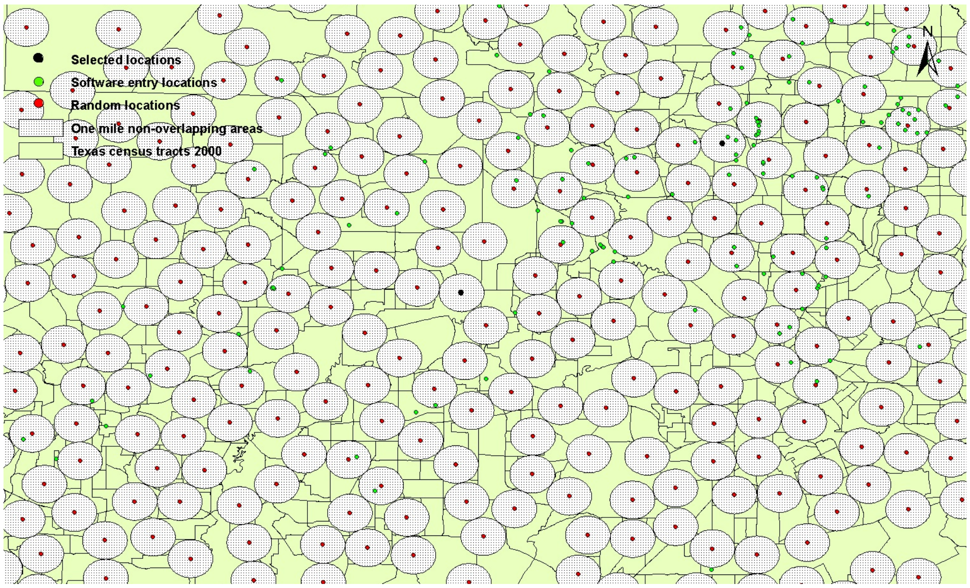


Panel A: 2000 Q2

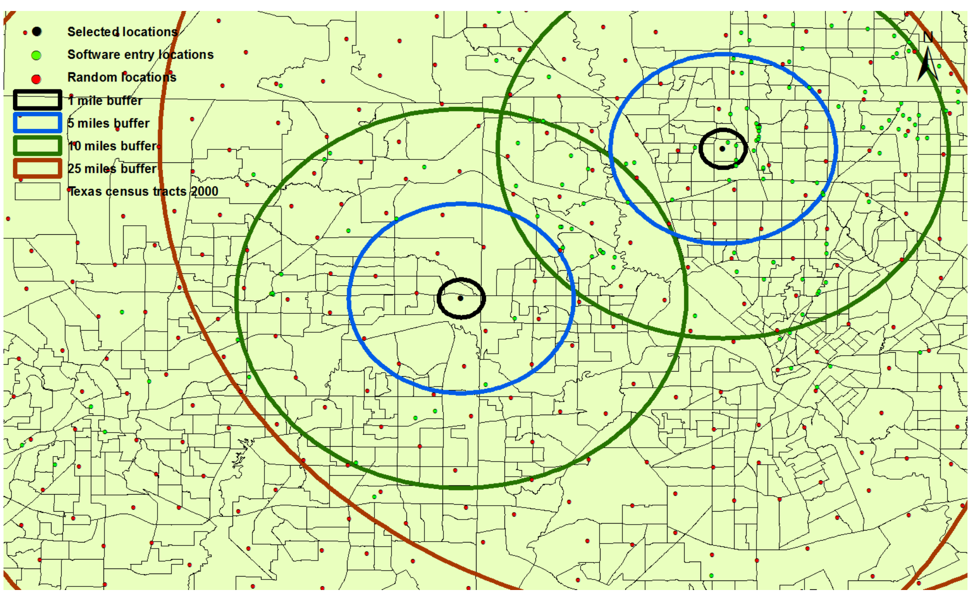


Panel B: 2006 Q4

Figure 5: Non-overlapping and selected locations in Dallas – Fort Worth



Panel A: Non-overlapping one mile rings



Panel B: Rings at two locations