**Lancaster University
Management School**

Spatial Persistence of Agglomeration in Software Publishing

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

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

Spatial Persistence of Agglomeration in Software Publishing*

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

June 8, 2018

Abstract

We estimate the effects of industrial localization on the spatial persistence of employment in the software industry, using micro-data from Texas for the 2000-2006 period. Locations with an initial concentration of software employment retain an excess number of employees, beyond that expected from job turnover and job persistence at the establishment level. This is not driven by differential establishment growth or survival, but it is due to (a) the retention by establishments in a location of jobs lost by other establishments in that location, and (b) the propensity of software establishments to enter in locations with prior software establishment presence. These findings are more consistent with labor channel effects than with human capital spillovers.

JEL Classification: R32, L86, R12.
Keywords: Agglomeration economies, labor pools, knowledge spillovers, firm growth, spatial effects.

1 Introduction

A voluminous literature has examined the ubiquity and sources of spatial aggregation at the industry level, and investigated the effects of agglomeration on technical change, firm growth, productivity, and other outcomes. To what extent, though, is industrial agglomeration self-sustaining? What are the drivers of its persistence? These questions are the focus of this paper.

Persistent agglomeration raises policy challenges for areas and localities that offer no advantage to any industry that benefits from agglomeration economies. If persistence is shown to be significant, it would suggest that such a locality could only succeed in becoming an activity hub through a concerted effort to attract simultaneous relocation or entry of multiple enterprises. It would also suggest that economic development efforts would be more successful if targeting some specific industries rather than being applied with no specific sectors in mind. However, if agglomeration were not persistent, a laissez faire approach would be appropriate. A policy aimed at facilitating industry clusters must identify sources of agglomeration economies which are responsible for the persistence of agglomeration. This is a challenging task, because agglomeration economies operate through many channels simultaneously.[1]

In this paper, we isolate the contributions of agglomeration spillovers on the spatial persistence of employment, from the effects arising from other factors. We accomplish this using data from the software publishing industry in Texas, which lends itself exceptionally well to the investigation of labor market and knowledge spillovers. The industry's output is intangible intellectual property that sells in the national and global markets. Thus, locational factors such as access to natural resources, local demand, proximate input suppliers, and transportation costs are not particularly relevant. Software publishing is a relatively new industry, hence historical factors should not have a significant impact on firm and industry location. 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. Moreover, given the 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

---

[1]These include the prevalence of specialized inputs and upstream service providers, labor pooling, and localized knowledge spillovers, as recognized since Marshall (1920).

programmers, the primary labor input, tend to be young and mobile.[2] Therefore, if labor-driven localization economies are present, their effects should be easy to detect. However, the absence of other plausible channels for such spillovers is a necessary but not a sufficient condition to measure the effects of labor channel localization economies. We must also control for localized economic activity that might otherwise result in spurious spatial association of software publishing. We do so using data on other business establishments at the address level, to account both for local economic development and also for the presence of infrastructure and zoning that is conducive for software publishing activity.[3]

Geographic clusters of software establishments are highly persistent, even when measured in locations that are only two miles across. Employment at the facility level exhibits very small persistence, but employment within the locations these facilities are situated is very persistent. Most of this persistence is driven by employment of entrants who locate in different addresses but very close proximity to existing firms. This geographic clustering of entrants around incumbents takes places despite the fact that the locations with incumbent firms cover a tiny portion of the Texas area with active business enterprises. These facts are highlighted in Figure 1. Total software employment in Texas has grown slowly during our sample period. Almost all employment has been in locations with active establishments at the start of our sample, with locations defined a very fine level of spatial dis-aggregation (1-mile radius circles; details provided in section 3.1). The preponderance of commercially active locations in Texas had no software publishers at the start of period and did not have any at the of our period either. Such persistence could have been explained by establishment-level persistence in employment. But this is not the case: employment in establishments active at the start of our period has declined substantially. The decline is steeper if one excludes job creation at the establishment level. The gap between these two lines represents jobs lost by an establishment that were "captured" by other establishments located in the same two-mile wide location.

Motivated by these facts, we formally estimate employment persistence at various level of aggregation, and explain it on the basis of a number of relevant economic factors. Within a 1-mile radius, we find that the elasticity of software employment at the end of 2006 with respect to employment at the start of 2000 is approximately $0.50 - 0.75$. These are very high numbers given that the majority of specific software jobs

---

[2]According to the Bureau of Labor Statistics, 51.5% of the employees in software publishing in 2017 were classified as computer occupation professionals.

[3]Texas has a mix of areas with high and low concentration of software establishments; thus our findings likely to be relevant to other states. We expect the labor market channel effects we identify to have external validity to other industries that employ high skilled labor, though in those industries other agglomeration effects, about which we have little to say, are also present.

that were present at the start of 2000 were lost. In fact, at the establishment level, we find the elasticity of software employment to be in the $0.30-0.35$ range. Our results also indicate that end-of-period employment is lower for establishments located in areas with existing software establishments, despite the absence of product side competition. Thus, the presence of software establishments reduces rather than enhances the success and growth prospects of other incumbents. We provide evidence that many of the jobs lost by a software establishment in a location were captured by other software establishments located no more than one mile away. This, however, is not sufficient to fully explain the high elasticity at the location level, leaving the location choices of entrants as the likely source of the additional jobs. To confirm this possibility, we investigate entry rates at different locations and follow the survival of these entrants as a function of location characteristics. After controlling for localized presence in a control group of other businesses, we find that software establishments tend to enter locations within one mile of other software establishments, thus reinforcing rather than diluting any initial location advantage. We find that the attractiveness of a location diminishes rapidly with distance and almost disappears after 5 miles. There are no comparable effects on exit at any distance threshold, as we show using survival analysis.

The high spatial persistence of employment in the software industry is driven by two components: the propensity of software establishments to locate in areas with pre-existing software establishments, and the ability of co-located establishments to act as a "sponge" for jobs lost by their neighbors. Spatial persistence is not driven by differential establishment growth rates in areas with substantial software industry presence or by differential establishment survival probabilities. Spatial persistence operates primarily via the labor market. Close proximity facilitates the transmission of information about employees of other firms and increases match value for prospective employers. It also reduces switching costs, as employees can retain living and commuting arrangements when switching employers, whether incumbent or entrant. These factors reduce the set-up costs of establishments and subsequent recruiting costs. Thus, entrants tend to locate near existing establishments and departing employees of a firm tend to locate a spin-off in the same vicinity.[4] Our finding that the presence of proximate software establishments (marginally) decreases employment growth at the establishment level is inconsistent, all else equal, with the presence of positive knowledge externalities. However, limited knowledge spillovers might be present and explain why the increased entry in

---

[4]City-wide policies aiming to attract software firms (rare in Texas to begin with) cannot drive our persistence findings given that our unit of analysis is very small.

high agglomeration locations of (some presumably more marginal) establishments does not depress average survival rates.

The findings of the importance of the labor pooling agglomeration effects for the spatial persistence of employment at the industry level are corroborated by studies that document the importance of labor pooling in generating agglomeration economies. Prominent among those are the contributions of Overman and Puga (2010) based on cross-industry analysis using UK data, Rosenthal and Strange (2001) who find in a cross-industry study that labor pooling effects are the dominant source of agglomeration beyond the zip code level, and Ellison, Glaeser and Kerr (2010) who measure the relative importance of a various sources of agglomeration, including labor pooling, also using cross-industry data. Our study differs from these in terms of the special features of the computer industry, and our focus on the link between agglomeration economies and spatial persistence of employment at the industry level.[5]

2   Conceptual Framework

We start by discussing knowledge and labor pool spillovers, and illustrate their impact on the persistence of agglomeration through a stylized framework. Since professional experience is embodied in the individual, worker mobility and direct personal and professional interaction are the primary channels of un-priced knowledge spillovers. Close geographic proximity facilitates these channels. For example, workers who happen to know each other can easily meet (e.g, for lunch, coffee, or even car-pooling) if their places of employment are proximate. Being within walking distance 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 although there may be "winners" and "losers" for any interaction. When one firm absorbs an idea from another, it might be to the detriment of the source firm if there is competition between them.[6]

The labor pool effect may have several expressions. A large pool of programmers within commuting distance allows a firm to expand more easily by poaching the employees of other firms, and reduces the set-up costs for new entrants. The overall effect on firm productivity and growth could be positive if

---

[5]Agglomeration economies 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).

[6]Combes (2000) notes that a greater number of similar firms in a locality increases the likelihood of knowledge spillovers due to the greater likelihood of closer matches between firms. Spillovers might conceivably even be negative in aggregate if they lead to free-riding. Firms might reduce experimentation with new ideas hoping to piggy back on ideas developed by other firms. This could possibly lead to a reduction of available knowledge.

employee-employer match quality improves by this practice. The labor channel can also operate through the founding of start-up firms by employees departing their current employer. As with knowledge spillovers, these localized effects may have both winners and losers with some successful firms benefiting to the detriment of their neighbors. In fact, as one industry executive confided to us, programmers employed by different firms compare work conditions and terms of employment when in social contact and switch employment when their current employer is not competitive.[7] Consistent with the labor pooling hypothesis, Freedman (2008) finds that spatial clustering facilitates localized worker mobility in the software publishing industry. Rosenthal and Strange (2003) report a quotation 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 transaction costs are probably negligible. The quotation has implications for the appropriate geographical area over which agglomeration externalities can be observed. Indeed, Rosenthal and Strange (2003), Wallsten (2001) and Arzaghi and Henderson (2008) note that localization externalities and knowledge spillovers attenuate rapidly within one mile.[8]

While both knowledge spillovers and labor pool effects may lead to a positive association between agglomeration and spatial persistence in employment, the effects differ subtly between these two sources of persistence. Labor market effects lead to the retention of jobs in a locale because the mere availability of labor attracts entry, even without improving the survival prospects of the entrants. Labor market effects can also improve firm growth and survival if thicker labor markets improve employee-firm matching and reduce ongoing costs of recruiting. In the absence of increased entry, increased growth and survival rates are unlikely to be driven by labor market effects.[9] By contrast, knowledge spillovers are more likely to lead to gains for existing firms and better survival prospects for entrants (given that software firms compete in a national market). Of course, by increasing growth and survival prospects, knowledge spillovers would also increase entry. However, in the absence of any growth and survival effects, increased entry is unlikely to be driven by knowledge spillovers. Labor market effects are a more persuasive explanation in this case. A

---

[7]Some Silicon Valley firms run company buses from San Francisco to their facilities to ease 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 firms.

[8]Within urban cores, a IT cluster may emerge in part because of social agglomeration effects, i.e., satisfaction obtained from socializing with individuals of similar backgrounds (see Hutton, 2004). Though of primary importance for dense inner cities, e.g., San Francisco, Manhattan, Vancouver, they may also play a minor role in Texas.

[9]Spatial persistence may be particularly high if the source of agglomeration economies is the labor pool channel. When an establishment receives a negative shock and lays off workers, it creates a pool of local jobseekers and reduces the recruiting costs of co-located establishments, facilitating their expansion. This tends to stabilize employment at the location.

very simple stylized model is developed in the Appendix that describes a single period equilibrium of this dynamic story.

To summarize, based on this framework, we interpret possible results as follows. If we observe no excess spatial persistence in software industry employment ("excess" being higher than expected by the employment persistence at the establishment level and the economic activity in that location), then neither labor market nor knowledge spillovers are important. If we do observe excess spatial persistence, then either knowledge spillovers, or labor market agglomeration economies, or both are present. We can distinguish the presence and relative importance of each as follows. If establishments in areas of high software agglomeration exhibit faster growth and higher survival rates, and if these areas have higher entry rates, then knowledge spillovers dominate, though weaker labor market effects may also be present. If establishments in areas of high software agglomeration exhibit faster growth and lower survival rates, but there are no positive entry effects, then only knowledge spillovers are present. If establishments in areas of high software agglomeration exhibit slower growth and lower survival rates, but there are positive entry effects, then only labor pool effects are present. Finally, if there are no growth and survival rate effects for establishments in areas of high software agglomeration, but there are positive entry effects, then labor pool effects dominate, but weaker knowledge spillovers may be present as well.

3 Spatial Sampling, Estimation Framework and Data

3.1 Data and Location Definition

The Quarterly Census of Employment and Wages (QCEW) compiled by the Texas Workforce Commission provides establishment data used in this study. Each record includes the specific location (address and latitude/longitude) of the establishment, business start-up date, and the relevant six-digit North American Industrial Classification System (NAICS) code. Separate establishments of a firm are reported in separate records. This data contains observations from the 1st quarter of 2000 through the 4th quarter of 2006.[10]

After restricting the analysis to the software publishing industry (NAICS code 511210), the sample contains 957 establishments corresponding to 877 unique firms (the vast majority of firms are single-establishment enterprises, and thus 'establishments' will be used inter-changeably with 'firms'). Average

---

[10]The authors obtained these data under an agreement of confidentiality and, therefore, disclosure of the actual data is subject to certain restrictions.

establishment size is around 35 employees. The number of software publishing establishments decreased from a high of 648 in Q4:2002 and Q1:2003 to a low of 581 in Q4:2006.[11] However, industry employment increased from 16,600 to 21,000 over the sample years, implying an increase in average establishment size.

A key feature of our analysis is to identify the geographic scale at which agglomeration effects are relevant. If they dissipate rapidly over space, observing geographies at the sub-county level is critical.[12] In principle, this would be accomplished by looking at county sub-divisions, such as census blocks, but these are of variable geographic extent and irregular shape. Ideally, we would like locations to be equally sized and defined in a purely spatial manner. Moreover, locations should only include plausible destinations for software establishments. At small scales, some areas do not offer basic conditions for the entry of software publishers. Including these areas is problematic in two ways. First, it is inconsistent with specifications that assume a positive expected number of software entrants. Second, it would result in a plethora of observations, only a tiny fraction of which would experience software establishment entry. In defining locations, we adopt the principle that human capital requirements can potentially be met either because required employees are already locally available or because they could be induced to move. 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 a software establishment cannot enter a location where these conditions are not satisfied.

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. The industries to which these other establishments belong are referred to as "control industries." There about 700 industries on this list (at the 6-digit level) constituting most of the employment in Texas.[13] We presume that any location that contains an establishment from these other control industries is one 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 is fungible across these industries and zoning laws will accommodate software publishing activities.[14] More-

---

[11] Fourteen software publishing establishments had PO Box as the official address, and their physical location could not be ascertained. These are not included in the above totals or in the subsequent analysis.

[12] Some of the variables are by necessity defined at the county level.

[13] A large number of industries share facilities with software establishments because many industrial and agricultural firms have offices located separately from production facilities. These offices can share a building with software firms. A weight scheme, described at the end of this section, mitigates issues when these industries are used to construct control variables.

[14] This does not imply that if a software establishment in some location shares a building with an advertising agency then

over, 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.

There are 580,375 unique control industry establishment locations (including establishments that entered or exited during our sample period).[15] We sort these establishments by longitude (west to east) and sequentially drop any that are closer than 2 miles to a previous establishment. We obtain a final sample of 9,299 establishments that form the center points of non-overlapping circle locations of one mile radius. These locations cover about 11 percent of Texas.[16] These locations are surrounded by concentric rings of 5, 10 and 25 mile radii.

When used in a panel, as in the entry analysis, the locations yield 251,073 observations (a balanced panel of 27 quarters; the first quarter is dropped because of lagging of some variables). In Table 1 Panel B we report the summary statistics for these observations. We note that nearly all software establishments are less than a mile from a control industry establishment, and thus these control industries effectively map out the large majority of potential locations.[17] However, after we drop locations to eliminate overlap, some spatial gaps are created, and thus the proportion of software establishments outside the final set of 1-mile circle locations rises to 28%. Figure 2, which shows the set of locations and software establishments in the Dallas area, is illustrative of this. The establishments in the "gaps" are lost from the left hand side of the analysis, but nearly all establishments are incorporated on the right hand side because they are within 25 miles of location centers. This is akin to explaining the behavior of a random sample of locations and establishments, but conditional on this sample, no errors-in-variables biases are introduced.[18] We use two measures of software firm activity within each geographical area: (i) the number of establishments, and (ii) the number of employees in these establishments.

The control industries are not only used to define locations, but also to control for the baseline propensity

every building housing an advertising agency could potentially house a software establishment. But most such buildings would be a potential location or would be close to other buildings that would be suitable locations.

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

[16]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. However, the covered area differs very little.

[17]Only 14 establishments are further than a mile from the nearest control firm facility. Most of these are one person operations, likely home-based.

[18]The only way to avoid gaps that potentially contain software establishments 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, it is geometrically challenging to define neighborhoods around square locations that contain all points no more than a specified distance from the edge and center of the location, and not contain points that are further than those pre-specified distances.

of software establishments to situate (or enter) there. Such control is necessary because locations with infrastructure that makes software establishment location possible differ in the extent to which they posses that infrastructure. Locations that contain a large number of control industry employees would, all things equal, be expected to contain more software establishments. Less developed locations, with a smaller number of control industry employees, are likely to contain fewer software establishments. Failure to control for this baseline might generate spurious clustering of entering establishments since a developing area attracts more entrants, including more software entrants. 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 economically less desirable).

We control for the baseline propensity of being situated in (or entering) a given location in two ways. The first way is to use the number of employees in the control industries as a control variable. The second way is to use the weighted number of employees as a control variable, 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 co-located with software facilities. We find the second approach more appealing and report the results using the weighted control (results are robust to using the unweighted number of employees).[19]

Key features of the data are provided in Table 1. Within one mile of an existing software establishment, there are, on average, approximately 10 other software establishments employing 394 employees. At distances between 1 and 5 miles, there are an additional 48 software establishments employing 1,857 employees, a substantial drop off in the density of software firm activity (recall area is proportional to the square of distance). The next 5 and the following 15 miles contain 46 and 72 software establishments employing 1,810 and 2,525 employees, respectively, corresponding to ever-larger density drop-offs. 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.[20] At small scales (up to 10 miles), the employment of software firms appears to drop off faster than that of the control firms.

Considering firm characteristics, we include a measure of the establishment's exposure to university

---

[19]Using collocation weights to construct the baseline activity level also accounts for possible spillovers from other firms, even though we expect these synergies to be small (see Arai et al, 2004).

[20]These averages are high because they are taken at the establishment-cross-quarter level. With most software establishments being located in the metropolitan areas (especially Dallas Metroplex and Austin), the typical urban density dominates the data.

R&D, which captures the possibility that knowledge spillovers are available from research universities.[21] This variable is based on the total federally funded research expenditures at the university located closest to the establishment, 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 (minimum 1 mile). We distinguish between research universities and colleges/junior colleges. 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 (Abramovsky, 2007).

We proxy the quality of the labor pool by local recreational and cultural amenities.[22] As Woodward et al. (2006) suggest, cultural and natural amenities are important in attracting and retaining skilled workers. To measure the local presence of these amenities, we compute the share of county employment in NAICS 71 (Arts, Entertainment, and Recreation), 721110 (hotels and motels), 722110 (full service restaurants), and 722410 (drinking places, alcoholic beverages) as reported in the QCEW data set. De Silva and McComb (2012) have used a similar measure. To account for factor costs, we use the average quarterly payroll of high-tech industries in the county.[23] The county unemployment rate for the final month in each quarter, as reported by the Texas Workforce Commission, proxies for economic conditions. In the bulk of the regression analysis, we exclude establishments that have only a single employee throughout the sample period.[24]

3.2 Some Facts about Spatial Persistence of Software Industry Employment

The software industry is dynamic. There is a large turnover of establishments and substantial changes in the scale of establishments over time. For example, 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 jobs that were initially in our dataset, only 7,015 (or about 41 percent) persisted in the same establishment until the end of the

---

[21]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). 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.

[22]Values of county variables are assigned on the basis of the co-ordinates of a location's centroid.

[23]For the few counties with no high-tech industries, we used the average wage of NAICS 3133 (Manufacturing), 51 (Information), 52 (Finance and Insurance), 54 (Professional, Scientific, and Technical Services), 55 (Management of Companies and Enterprises), 61 (Educational Services), and 62 (Health Care and Social Assistance).

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

period.[25] The 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 period, 431 new establishments entered, almost all in different buildings than exiting establishments.[26] The jobs created by the entrants and the jobs added by growing incumbents raised total employment in the industry to about 21,000 (an increase of 24%). This means that 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. Given that this industry does not rely on specialized infrastructure and has non-localized demand for its product, and given that entrants typically choose different addresses than incumbent or exiting establishments, there is a potential for the spatial distribution of this industry to be completely transformed.

This turns out to not 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 list 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. While these figures indicate that concentration is broadly constant 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 a pictorial examination of a few representative areas. One cannot easily plot employment into space, but 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 where software publishers could be located. Note that software publishers are not uniformly distributed in this space, but are rather concentrated in certain areas. Moreover, areas of concentration remain stable despite entry and exit: an area with prior concentration of software publishers retains that concentration to the end of our sample period. A final observation relates to newly developed locations. Some areas were commercially undeveloped in the second quarter of 2000 and got developed by the end of 2006. But these areas did not attract a cluster of

---

[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.

[26]Establishments are identified by their address. In the rare case of a firm merger, constituent establishments remain distinct.

software establishments, and isolated software firm entry in them is rare (only 5% of software entrants locate more than 5 miles from an existing software facility).

Some additional evidence at the micro level is obtained by computing the number of software employees in each of the 1-mile radius locations and then seeing 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 at 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 industry at the establishment level.

4   Econometric Analysis and Results

4.1   Spatial Persistence in Employment and Its Components

A more formal analysis would perform the following conceptual exercise. Suppose we could exogenously increase the software employment in a particular location at the start of the period by one percent (e.g., by increasing the order flow of the establishments situated there). What would then be the percentage increase in software employment in that location at the end of the sample period?[27] Clearly, we cannot observe the impact of exogenously created software jobs being in a location. Rather, we observe locations that differ in the level of initial software employment and other characteristics. Using locational controls based on industries that are typically co-located, and noting that the software industry has minimal infrastructure demands and a non-physical product with a national market, we ensure that the initial variation in employment comes close to being exogenous in the statistical sense.[28]

To better understand the content of the question we investigate, we observe that end-of-period employ-

---

[27]The same hypothetical question could be expressed in terms of employment levels rather than percentages, in which case it would correspond to analysis using the number of employees rather than their log.

[28]Initial variation in employment need not be random, but only uncorrelated with unobserved shocks to end-of-sample period employment.

ment in a location is a function of jobs lost and accrued. In all of this analysis, a job in an establishment is retained regardless of employee turnover. For the moment, suppose that every job in any establishment 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 is 1. However, such an extreme degree of persistence is unlikely. Entrants 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 literature on Gibrat's law) and hence final employment in a location is not expected to be proportional to initial employment in that location. As a result, the landscape of the software industry activity would shift over time. Some areas with high concentrations would "revert to the mean" or 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 industry activity grew systematically faster than establishments located in areas with a low concentration. The degree of spatial persistence of employment would balance these factors. In the absence of agglomeration economies, we would expect spatial persistence to be driven solely by the inertia of jobs at the establishment level.[29] With agglomeration economies, we would expect it to be higher than this value. Indeed, the observed fraction of initial jobs that remain with the same employer is so low as to be consistent with even a *negative* elasticity.[30] Thus, even a demonstration that the elasticity of final employment to initial employment is positive has empirical significance.

We now describe the econometric framework through which we investigate the extent of spatial persistence. Our spatial unit of econometric analysis is the one mile radius locations described in section 3.1. 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 establishments in the location at the start of the sample period, the number of software establishments in concentric 1-5, 5-10 and 10-25 mile rings and the number of employees of these establishments. Though the

---

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

[30]Consider two establishments (or locations), one with initial employment of 100, the other with initial employment of 200. Let the final period employment levels be flipped. Then, the number of retained jobs is 100 for each establishment, or a total of 200 (67 percent) for the industry. But the elasticity of final employment with respect to initial employment is negative!

1-mile radius locations are not overlapping, the surrounding rings are.

County effects for the five counties with major employment in this industry are included in some specifications.[31] For these regressions, the effect of initial conditions is identified from the within-county distribution of software publishing activity. 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 establishments and employees in the concentric rings surrounding a location. These latter variables are co-determined to some extent (though the stronger influence probably goes from the larger outer rings to the center). More formally, our most general regression model is given by

$$Y_{l,T} = D'_{l,t=1}\delta_0 + R'_{l,t=1}\rho_0 + C'_{c_l,t=1}\varphi_0 + D'_{l,T-1}\delta_1 + R'_{l,T-1}\rho_1 + C'_{c_l,T-1}\varphi_1 + \eta_l \tag{1}$$

where $Y_{l,T}$ is the location l's quarterly software employment in the final quarter, T (Q4:2006), or its natural log. 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. We base all inferences on robust standard errors for parameters and marginal effects.

Employment has a long right tail, and many of the explanatory factors are expected to act synergistically, rather than in a purely additive fashion. Thus, an analysis using log-employment as the dependent variable is more appropriate, though we report results using both specifications. The lowest possible employment level is 1, so $Y_{l,T}$ is censored and equation (1) is estimated using censored regression methods.[32] One quarter of the observations are censored, which might seem high. However, about 70 percent of the locations initially have only one establishment and these locations form the bulk of those that are censored (i.e., the sole establishment in that location exited and was not replaced). Given that half of the incumbent establishments fail, a 25% censoring rate implies that in many locations with only one establishment, at least one other establishment was attracted prior to the exit of the incumbent. This suggests a propensity of entrants to locate close to incumbents, which is formally demonstrated in section 4.2.

---

[31]Using country fixed effects for all counties effectively "dummies out" most locations situated in those counties, without major changes in the findings.

[32]The censoring threshold in the log linear specification is 0, while in the linear specification it is 1. We distinguish between locations with one employee and censored locations with no software employment.

Table 2 presents the censored regression results using log employment as the dependent variable, moving from the simplest to the more complicated specifications. Since our focus is on the marginal effect of initial employment on the expected value of final employment, this table (and all tables with employment as the dependent variable) report marginal effects and the associated robust standard errors. Table 3 presents the results using the same sets of regressors but with linear rather than log specifications.[33]

We discuss the results of Tables 2 and 3 together. In the simplest specification (model 1), no covariates are used except for differential intercepts for major counties. The point estimates suggest that a one percent increase in initial employment translates into a 0.66-0.75 percent increase in final period employment. This indicates a large spatial persistence in line with the informal evidence described earlier. The underlying parameter estimates are higher still, and not significantly different from 1. 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 are jointly significant, employment at any distance up to 25 miles is not. Controlling for localized employment and growth in the control industries (model 3) also leads to only minor reductions in 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. Establishments in a location proximate to areas with many software employees can obtain a ready supply of workers, but if those proximate areas are growing themselves, they can siphon workers away from that location.

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)

---

[33]To make a valid comparison between the marginal effects in the two tables, those of the regressions in Table 3 are weighted by the initial number of employees in each location, i.e., they are the sample average of $\frac{Y_{l,t=0}}{Y_{t=0}} \frac{\partial E[Y_{l,T}]}{\partial x_l}$ where $Y_{t=0}$ is the initial software employment in Texas and x_l is any regressor.

and those arising from an increase in employment (holding the number of facilities constant). Both appear generally significant. In the log specification (where the coefficients are both elasticities) the estimated coefficient on the number of establishments is larger (though less precisely estimated). Remarkably, a "replication" of the facilities in a location, seems to double end of period employment in that location, i.e., there is nearly full persistence.[34] Part of the explanation why the number of initially active establishments is associated with higher terminal employment, which we have verified in our data, is that it increases the probability that some of these establishments grow substantially and reduces the possibility that all of them fail. Estimates at higher distances are zero, except those for the 5 to 10 mile range (but where the number of employees and number of establishments enter with opposite sign in the log specification). When including final employment and number of establishments (model 6), results are the same as those for the conceptually similar model 4.

In light of earlier discussion, 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 2 and 3 with the counterpart specifications estimated at the establishment, rather than the location level.[35]

Examining model 1 of Tables 4 and 5, which is the direct counterpart of model 1 of Tables 2 and 3, the estimate of the coefficient of initial establishment employment on final period establishment employment is much smaller than those obtained at the location level.[36] Increasing initial employment in an establishment by one percent leads to only 0.3 to 0.4 percent increase in final employment. Models 2, 3 and 4 are progressively more inclusive specifications and confirm this pattern. These results corroborate the conclusion that establishment level employment persistence is much lower than location level persistence.[37]

Conceptually, the difference between establishment-level and location-level employment persistence consists of three components: (i) faster growth rates of establishments in high concentration locations, (ii) the

---

[34]Replicating the facilities results in doubling the number of establishments and doubling the total employment. The implied end-of-period effect from Table 2, column 5, is $0.483 + 0.467 = 0.95$ while from column 6, it is $0.496 + 0.606 = 1.102$.

[35]Note that 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.

[36]A coefficient of 1 would imply Gibrat's law. Most of the literature investigating Gibrat's law has 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).

[37]These estimates are not driven by small establishments that contribute little to aggregate employment; re-estimating Tables 4 and 5 using initial employment weights yields very similar results.

capture of jobs lost to an establishment by others in that same 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-location, yielding higher employment growth among establishments present in a location with many other co-located establishments. 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.

We investigate this key question using establishment-level regressions by re-estimated models 1-4 of Tables 4 and 5 after adding the initial employment by co-located establishments as a regressor (plus 1, when taking the log). The results, reported in columns 5-8 of Tables 4 and 5, 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.[38] With the addition of both number and employment of other establishments in the specification, the sign on the number of other establishments is positive (though significant only in logs). However, the coefficient on the number of employees of those establishments becomes even more negative. This means, then, that the effect of initial employment on final employment does not arise from a positive effect that initial employment has on the growth of the existing establishments. It rather 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 examine whether the local capture of jobs lost by an establishment is indeed a contributor to local employment persistence. This is done by re-estimating the regressions in Tables 2 and 3 using as dependent variable the final employment in a location's establishments that were present at the start of the sample period (incumbents). As reported in Tables 6 and 7, the coefficients on initial employment are higher than those of Tables 4 and 5. This indicates 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.[39]

---

[38]It is conceptually possible that there are positive synergies, and that these attract entrants of lower intrinsic productivity to high density locations, resulting in this negative association. We have no direct measures of establishment productivity to test this proposition, but the null results of the duration analysis in section 4.2 suggest these effects are not material.

[39]For an illustrative 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 ones. 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

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 2 and 3. For example, Model 1 of Table 2 yields an elasticity of 0.656, while the elasticity is 0.528 when incumbent establishment employment is the dependent variable (Table 6), and 0.366 at the establishment level (Table 4). For the linear version of this model, the elasticity measured by the weighted marginal effect is 0.753 when employment of all establishments is the dependent variable (Table 3) versus 0.539 when incumbent establishment employment is the dependent variable (Table 7). This compares to 0.283 at the establishment level (Table 5). Similar results are obtained when comparing the other models in these tables. In other words, the employment persistence in a locality is only partially driven by incumbent establishments capturing jobs lost to other incumbents. Some of the persistence must be driven by disproportionate entry rates into locations with greater initial presence of software publishers.

4.2 Establishment Entry and Survival

In this section, we examine 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 activity in a location influences the entry rate of software firms in that location. Since other factors are associated with entry rates, we must carefully formulate the econometric design. Even under the null hypothesis that localization economies do not influence entry probabilities, the expected number of entrants is not uniform across all one-mile radius locations. 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. Of these characteristics, the number of employees in other related industries, weighted by each industry's spatial association with software publishing firms, is the most important control variable. This is not a causal relationship, but serves as a summary statistic for the baseline propensity of software firms to enter that location.

An important difference between the employment persistence analysis and the entry analysis is that the latter takes advantage of the time variation in our data. We estimate the number of establishments entering a location in a particular quarter as a function of the location characteristics in the preceding quarter. Doing

---

are balanced by gains in the small establishment in that same location.

18

so increases the variation we can exploit in our sample. In the typical entry case, there is a single software entrant in a given location in a given quarter, while the maximum number of entering software establishments is two (except for one occurrence of 5). Given this, we estimate the entry models using an ordered probit. Our dependent variable is the number of software publishing start-ups $y_{l,t}$ 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 and 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 characterized by long tails are in logs. When taking the log of the number of establishments or employees, the value of 1 is added to ensure well defined regressor values for a location with no establishments. In some specifications we have included county effects α_{c_l} for the top 5 counties with the most software publishers.[40] The random disturbance ϵ_{lt} has a standard normal distribution.

The results are reported in Table 8. Localized own-industry density has a strong positive effect on entry probabilities regardless of whether the number of software establishments or the associated number of employees is used as a measure of density. The effect appears to be stronger for distances of less than one mile, somewhat important for intermediate distances (1-5 miles) and marginal or absent for distances greater than 5 miles (especially for more comprehensive specifications). The use of county fixed effects tends to weaken the association between pre-existing software establishments 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 using similar infrastructure (Other Industry Employees) is positive and strongly significant. The number of software establishments seems equally important for the location decision of potential entrants as the total number of employees of those establishments (measured

---

[40]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 repeat entries into any particular location are rare. Moreover, adding dummies for each county results in the loss of many observations.

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 establishments as the measure of software publisher presence having a slight edge.

Finally, when both the number of establishments 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) while junior college spillovers and local unemployment rates are negatively associated with entry.[41]

Despite the small entry counts, we also estimate the entry process using the Poisson model in order to investigate the sensitivity of the results to the econometric specification. The estimation results are not reported for brevity, but are summarized here. Though the parameter values of the Poisson and ordered probit models are not directly comparable with each other because of differences in scaling, 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 mostly similar. However, localized own-industry density effects now appear slightly stronger for the 1 to 5 mile distance, as for distances of up to one mile, but only when employment levels are the measure of industry activity.

We next investigate whether establishments entering areas with high prior software firm concentration have higher survival rates. We estimate a number of different specifications of Weibull duration models, using variable sets that parallel those of entry models (though variable values correspond to the date of establishment entry). We find no statistically significant association between the number of other software establishments or their employment and the exit rates of an establishment. These results are corroborated by a discrete choice model of the exit decision (both sets of results are available upon request).

---

[41]We have also estimated these models using ordered logit, obtaining qualitatively similar estimates for the agglomeration variables. 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.

5 Interpretation of Results and Concluding Remarks

Software industry employment at the highly localized level (one mile radius) exhibits excess persistence, i.e., greater persistence than implied by the employment persistence of individual establishments. This persistence is not driven by the faster employment growth of establishments in areas with a high density of software publishing activity. Indeed, incumbent establishments seem to exhibit slower growth when located in close proximity to other establishments. Moreover, the persistence is only partially driven by the fact that a job "lost" by an establishment is "captured" by another in the same location. Rather, the presence of software firms increases the propensity of other software firms to enter within close proximity. But establishments entering localities with prior software firm presence do not experience differential survival rates than establishments entering localities with no prior software firm activity.

This combination of findings suggests that a prior concentration of software establishments in a locality lowers the entry costs of other software firms in that same locality, but that post-entry profitability 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 firms preferring to locate in a particular location because recruiting is easier.

The competing explanation, that entrants locate in an area because of synergies or direct human capital spillovers from incumbents, does not fit the facts. Our conceptual framework (see Panel A of Figure A-1) reveals that such spillovers should make the average entrant into that location more successful compared to entrants elsewhere and should also benefit incumbents. There is no evidence of this at the 1 mile range and, therefore, human capital spillovers cannot be the only source of spatial persistence. However, some human capital spillovers might be present in conjunction with labor channel effects. These human capital spillovers would reinforce the positive effects of labor pool effects on post-entry profits in high spatial agglomeration locations. Even though lower entry costs in those locations mean that software establishments co-locating with other software establishments have on average weaker business prospects, the upward shift in the payoff function would leave average exit rates unchanged (and employment growth of incumbent establishments only marginal lower).

We find no strong relationship between end-of-period software employment in a location and software employment at intermediate distances (1 to 10 miles) from that location. However, there is still a positive,

but diminished, association between entry rates and prior software publishing activity. There continues to be no association between exit rates and software publishing activity. In other words, entry costs in a location appear lower if there is a nearby labor pool, but the surrounding pool does not enhance the productivity of software establishments in that location, which supports the labor pool interpretation we suggested earlier.

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

Our findings suggest that spatial persistence is primarily driven by labor pool agglomeration effects. Productivity spillovers from human capital and knowledge transmission are of subsidiary importance, or absent. Firms are attracted to locations because of the existing labor force, not because entering there gives them better prospects. In this industry, jobs go where the workers are, rather than the other way around.

Appendix: A Stylized Model

Consider all potential establishments in a location l and rank them on the basis of their expected present value gross of entry costs. Let this present value be given by $V(p) = \theta(p) + \kappa z_l + \lambda \theta(p) z_l$, where p is the percentile ranking of an establishment, $\theta(\cdot)$ is an increasing function, z_l is a measure of value-enhancing spillovers in location l, and κ and λ are positive constants. Let the fixed costs of entry of any establishment be $FC = \gamma_0 + \gamma_1 x_l$, where x_l is a measure of the entry barriers in a location and γ_0 and γ_1 are constants.

For any location, l, the establishments with a gross present value exceeding entry costs enters and operates; the rest stay out. The fraction of potential establishments that enter, i.e., the entry threshold percentile, is given by $\hat{p}_l = \theta^{-1}\left(\frac{\gamma_0 + \gamma_1 x_l - \kappa z_l}{1 + \lambda z_l}\right)$. This is decreasing in any value-enhancing spillovers and any location characteristic that reduces entry costs. In this framework spillovers do not necessarily increase the present value of all establishments, but they do shift the distribution of present value so that a distribution with higher first-order spillovers dominates one with lower spillovers.[42]

Figure A-1, panel A, illustrates within this framework how entry and establishment performance differ across two locations, one with low and the other with high knowledge spillovers. The horizontal axis plots the percentiles of potential entrants, and the vertical axis plots the entry costs and the gross present value. The upward sloping lines are the relationship between establishment present value and establishment percentile rank for low and high spillover locations. For this figure, $\theta(\cdot)$ is assumed to be linear. The line for the high knowledge spillovers location is above the line for the low knowledge spillover location, indicating that

---

[42]For simplicity, fixed costs are the same for all establishments, but an extension that relaxes this is discussed in footnote 44.

knowledge spillovers are value enhancing, i.e., increase the value of z_l. The high z_l line is steeper, as $\lambda > 0$, suggesting a stronger positive effect at the high end of value distribution. The fixed costs of establishment entry are the same for high and low knowledge spillover areas, consistent with the premise that knowledge spillovers do not affect entry costs. The marginal entrant in the low and high spillover location is $\hat{p}^{**}$ and $\hat{p}^*$, respectively. High spillover locations attract more establishments. Moreover, the average performance (value) of these establishments is higher. With the distribution of establishments over percentiles being uniform by construction, the mean expected value of the operating establishments is increasing in z_l.[43] The continuation value and (future) employment of an establishment evolves over time, but is positively correlated with its initial expected value. Thus, locations with high values of z_l are expected to have more and faster growing establishments. When an establishment's continuation value drops below zero, it exits. With the distribution of establishment expected present values being higher in locations with high values of z_l, exit rates in those locations should be lower.

Panel B of Figure A-1 illustrates how entry and establishment performance outcomes differ across two locations, one with low and the other with high labor market agglomeration economies. Agglomeration effects arising from the labor channel reduce the cost of entry. They can also reduce ongoing recruiting costs and thus shift up the distribution of gross present value of establishments. In our notation, labor market agglomeration economies decrease x_l and increase z_l. The second effect is expected to be smaller. Whether or not this is the case, the important point is that the labor channel affects both set-up costs and also future performance. Panel B of Figure A-1 shows that the average present value of entering establishments in locations with high labor agglomeration may be lower, even though they attract more establishments (as demonstrated in footnote 43, the expected value conditional on entry falls as x_l decreases, but rises with z_l). In fact, we believe that it is lower, since the entry effect is expected to dominate, but this depends on the relative size of the shifts of the two lines and the extent to which $\theta(\cdot)$ departs from linearity (and in which way).[44] Following the discussion in the preceding paragraph, the expected value of an establishment affects its future employment level and exit probability.[45] In markets with high labor agglomeration economies, then, establishment growth rates may be lower and exit probabilities higher than in locations with low agglomeration economies.

These two figures are both static representations and describe entry in a single period for a specific level of spatial agglomeration. Spatial agglomeration persists over time but persistence is not perfect, as random variation in establishment outcomes weakens the initial advantage of some locations and strengthens that of others.

---

[43]$E[V(p)|p > \hat{p}_l] = (1 + \lambda z_l)E[p|p > \hat{p}_l] + \kappa z_l$. Observing that $\hat{p}_l = (\gamma_0 + \gamma_1 x_l - \kappa z_l)/(1 + \lambda z_l)$, and that by definition p is distributed uniformly, we obtain $E[p|p > \hat{p}_l] = 1/2 + (\gamma_0 + \gamma_1 x_l - \kappa z_l)/(2(1 + \lambda z_l))$. Substituting in the expression for the establishment's expected value and simplifying, we get $E[V(p)|p > \hat{p}_l] = (1 + \lambda z_l + \gamma_0 + \gamma_1 x_l + \kappa z_l)/2$ which is increasing in z_l.

[44]Heterogeneous entry costs do not alter any of this discussion as long as establishments with high post-entry value enter "ahead" of those with low post-entry value. A more substantial modification of this framework, under which establishments 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.

[45]We note parenthetically that 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 establishment employment growth, while affecting survival probabilities.

23

References

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

[2] Arai, Y., H. Nakamura, H. Sato, T. Nakazawa, T. Musha, and K. Sugizaki (2004). "Multimedia and Internet Business Clusters in Central Tokyo," *Urban Geography*, 25: 483-500.

[3] Arzaghi, M. and J. V. Henderson (2008). "Networking off Madison Avenue," *Review of Economic Studies*, 75: 1011-1038.

[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 Start-ups 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., 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.

[10] Ellison, G., E. L. Glaeser and W. R. Kerr (2010). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, 100: 1195-1213.

[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] Hutton, T. A. (2004). "The New Economy of the Inner City," *Cities*, 21: 89-108.

[18] Marshall, A. (1920). *Principles of Economics*, London: MacMillan.

[19] Overman, H. G. and D. Puga (2010). "Labor Pooling as a Source of Agglomeration: An Empirical Investigation," in Glaeser, E. L., editor, *Agglomeration Economics*, chapter 4, pages 133-150, The University of Chicago Press.

[20] Rosenthal, S. S. and W. C. Strange (2001). "The Determinants of Agglomeration," *Journal of Urban Economics*, 50: 191-229.

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

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

[23] 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.

[24] 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.

[25] 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 establishments

Panel A: Full sample of establishments

	All	Incumbents	Entrants[a]
Unique number of firms	877	506	371
Number of establishments per firm	1.228	1.106	1.552
	(1.041)	(0.583)	(1.707)
Unique number of establishments	957	526	431
Number of firms with one establishment	857	497	360
Average size (employees) of establishments	34.955	31.071	45.256
	(194.517)	(113.474)	(322.322)
Own quarterly wage ($) per establishment	20,305.82	19,089.65	23,530.49
	(25,229.45)	(20,133.96)	(35,147.05)
Age (in months) per establishment	105.983	129.217	44.379
	(85.166)	(83.681)	(51.886)
Average time in the sample (quarters)	11.7000	13.038	8.150
	(7.6111)	(7.758)	(5.880)
University spillover ($)	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 software employment:	1.478	65.870	68.064
within a mile	(34.464)	(224.785)	(241.105)
Average software employment:	33.078	661.003	859.977
1 - 5 miles	(284.603)	(1,241.156)	(1,400.822)
Average other industry employment:	646.724	10,718.010	11,798.320
within a mile	(3,828.118)	(19,788.580)	(21,665.510)
Average other industry employment:	11,985.390	138,240.400	144,071.500
1 - 5 miles	(42,743.660)	(141,583.400)	(145,221.700)
Average number of software entrants	0.001	0.033	0.063
	(0.036)	(0.195)	(0.259)
Average size (employees) of other	11.281	24.882	28.495
establishments	(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: Censored log–linear regression results for software employment growth at a location

Variable	Log(*final employment*)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log initial software employment:	0.656***	0.599***	0.502***	0.530***	0.483***	0.496***
within a mile	(0.057)	(0.065)	(0.072)	(0.068)	(0.087)	(0.082)
Log initial software employment:		0.062	-0.010	0.162*	0.006	0.163
1 - 5 miles		(0.052)	(0.054)	(0.090)	(0.121)	(0.135)
Log initial software employment:		-0.141	-0.164*	-0.116	-0.545**	-0.359
5 - 10 miles		(0.088)	(0.085)	(0.163)	(0.214)	(0.230)
Log initial software employment:		0.116	0.101	0.428***	-0.044	0.221
10 - 25 miles		(0.076)	(0.075)	(0.162)	(0.130)	(0.191)
Log initial weighted other industry			0.118	0.141	0.126	0.187
employment: within a mile			(0.172)	(0.188)	(0.182)	(0.194)
Log initial number of software					0.467	0.606*
establishments: within a mile					(0.341)	(0.332)
Log initial number of software					-0.142	0.258
establishments: 1 - 5 miles					(0.298)	(0.504)
Log initial number of software					0.788**	1.058*
establishments: 5 - 10 miles					(0.401)	(0.543)
Log initial number of software					0.238	0.972*
establishmentss: 10 - 25 miles					(0.317)	(0.578)
Log final number of other software						-0.226
establishments: 1 - 5 miles						(0.439)
Log final number of other software						-0.520
establishments: 5 - 10 miles						(0.632)
Log final number of other software						-0.303
establishments: $10 - 25$ miles						(0.608)
Log final software employment:				-0.218**		-0.264*
$1 - 5$ miles				(0.087)		(0.139)
Log final software employment:				-0.055		-0.123
$5 - 10$ miles				(0.147)		(0.204)
Log final software employment:				-0.397**		-0.455**
$10 - 25$ miles				(0.168)		(0.221)
Log final weighted number of other			0.125	0.138	0.073	0.046
industry employment: within a mile			(0.189)	(0.204)	(0.207)	(0.211)
Initial period variables		Yes**	Yes**	Yes**	Yes**	Yes**
Final period variables				Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Number of obs.	201	201	201	201	201	201
Number of uncensored obs.	150	150	150	150	150	150
Log likelihood	-350.207	-337.907	-331.673	-325.971	-326.747	-320.076

Table reports marginal effects; associated robust standard errors are in parentheses. *** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percentlevel and * denotes statistical significance at the 10 percent level. The dependent variable is the log of total employment in non-overlapping locations in the last period. Initial and final period variables include spillovers, unemployment rate, and county high-tech wages. Final variables are lagged by one period. See text for details.

Table 3: Censored linear regression results for software employment growth at a location

Variable	Final employment					
	(1)	(2)	(3)	(4)	(5)	(6)
Initial software employment:	0.753***	0.739***	0.696***	0.693***	0.645***	0.617***
within a mile	(0.199)	(0.186)	(0.161)	(0.156)	(0.137)	(0.131)
Initial software employment:		0.011	0.004	0.132*	-0.030	0.117*
1 - 5 miles		(0.017)	(0.016)	(0.068)	(0.041)	(0.061)
Initial software employment:		-0.024	-0.024	0.065*	-0.048	0.059
5 - 10 miles		(0.022)	(0.022)	(0.037)	(0.036)	(0.042)
Initial software employment:		0.031	0.031	0.096	0.019	0.083**
10 - 25 miles		(0.029)	(0.030)	(0.058)	(0.016)	(0.042)
Initial weighted other industry			-0.026	-0.008	-0.003	0.036
employment: within a mile			(0.062)	(0.068)	(0.062)	(0.073)
Initial number of software					33.519**	44.492***
establishments: within a mile					(14.200)	(17.089)
Initial number of software					1.760	6.003
establishments: 1 - 5 miles					(2.477)	(7.997)
Initial number of software					0.815	7.126*
establishments: 5 - 10 miles					(1.800)	(4.124)
Initial number of software					1.173	-1.137
establishments: 10 - 25 miles					(2.132)	(3.626)
Final number of other software						0.796
establishments: 1 - 5 miles						(6.219)
Final number of other software						-5.417
establishments: 5 - 10 miles						(4.186)
Final number of other software						6.102
establishments: $10 - 25$ miles						(5.057)
Final software employment:				-0.119*		-0.206**
$1 - 5$ miles				(0.064)		(0.103)
Final software employment:				-0.079*		-0.110*
$5 - 10$ miles				(0.043)		(0.066)
Final software employment:				-0.058*		-0.102*
$10 - 25$ miles				(0.035)		(0.057)
Weighted other industry			0.057	0.055	0.014	-0.013
employment: within a mile			(0.036)	(0.038)	(0.040)	(0.051)
Initial period variables		Yes**	Yes**	Yes**	Yes**	Yes**
Final period variables				Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Number of obs.	201	201	201	201	201	201
Number of uncensored obs.	150	150	150	150	150	150
Log likelihood	-1,081.353	-1,074.429	-1,072.3045	-1,066.539	-1,069.064	-1,055.148

Table reports weighted marginal effects; associated robust standard errors are in parentheses. *** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percentlevel and * denotes statistical significance at the 10 percent level. The dependent variable is the total employment in non-overlapping locations in the last period. Initial and final period variables include spillovers, unemployment rate, and county high-tech wages. Final variables are lagged by one period. See text for details.

Table 4: Censored log–linear regression results for software establishment–level employment growth

Variable	Establishment-level log(*final employment*)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of initial establishment-level	0.366***	0.363***	0.347***	0.351***	0.402***	0.399***	0.391***	0.408***
employment	(0.041)	(0.042)	(0.043)	(0.043)	(0.046)	(0.046)	(0.046)	(0.049)
Log initial other software					-0.069*	-0.077*	-0.124***	-0.184**
employment: within a mile					(0.036)	(0.042)	(0.046)	(0.073)
Log initial other software		-0.053	-0.076*	-0.016		-0.022	-0.045	-0.080
employment: 1 - 5 miles		(0.038)	(0.041)	(0.101)		(0.042)	(0.043)	(0.104)
Log initial other software		0.084	0.087	0.119		0.074	0.073	0.075
employment: 5 - 10 miles		(0.056)	(0.057)	(0.134)		(0.056)	(0.056)	(0.135)
Log initial other software		-0.058	-0.048	-0.037		-0.065	-0.050	-0.072
employment: 10 - 25 miles		(0.053)	(0.053)	(0.174)		(0.052)	(0.053)	(0.173)
Log initial weighted other industry			0.057	0.079*			0.105**	0.103**
employment: within a mile			(0.039)	(0.042)			(0.043)	(0.043)
Log initial number of other software				-0.128				0.151
establishments: within a mile				(0.107)				(0.154)
Log initial number of other software				-0.069				0.056
establishments: 1 - 5 miles				(0.210)				(0.215)
Log initial number of other software				-0.076				-0.022
establishments: 5 - 10 miles				(0.266)				(0.266)
Log initial number of other software				-0.014				0.049
establishments: 10 - 25 miles				(0.283)				(0.283)
Initial period 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
Number of uncensored obs.	268	268	268	268	268	268	268	268
Log likelihood	-792.693	-789.586	-788.534	-787.477	-790.567	-787.887	-784.896	-784.313

Table reports marginal effects; associated robust standard errors are in parentheses. *** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percentlevel and * denotes statistical significance at the 10 percent level. The dependent variable is the log of total employment in an establishment in the last period. Initial period variables include spillovers, unemployment rate, and county high-tech wages.

Table 5: Censored linear regression results for software establishment–level employment growth

Variable	Establishment-level employment							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Initial establishment-level	0.283***	0.285***	0.284***	0.289***	0.290***	0.293***	0.294***	0.297***
employment	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)	(0.025)	(0.025)	(0.025)
Initial other software					-0.008	-0.010	-0.013*	-0.015
employment: within a mile					(0.007)	(0.007)	(0.008)	(0.010)
Initial other software		0.001	0.001	0.009		0.002	0.001	0.003
employment: 1 - 5 miles		(0.003)	(0.003)	(0.006)		(0.003)	(0.003)	(0.007)
Initial other software		-0.001	-0.001	0.001		-0.001	-0.001	-0.002
employment: 5 - 10 miles		(0.003)	(0.003)	(0.006)		(0.003)	(0.003)	(0.006)
Initial other software		0.000	0.000	-0.003		-0.000	-0.000	-0.005
employment: 10 - 25 miles		(0.002)	(0.002)	(0.006)		(0.002)	(0.002)	(0.006)
Initial weighted other industry			0.003	0.003			0.004	0.004
employment: within a mile			(0.003)	(0.003)			(0.003)	(0.003)
Initial number of other software				0.091				0.530
establishments: within a mile				(0.502)				(0.587)
Initial number of other software				-0.461				-0.207
establishments: 1 - 5 miles				(0.288)				(0.338)
Initial number of other software				-0.068				0.063
establishments: 5 - 10 miles				(0.279)				(0.293)
Initial number of other software				0.153				0.207
establishments: 10 - 25 miles				(0.222)				(0.225)
Initial period 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
Number of uncensored obs.	268	268	268	268	268	268	268	268
Log likelihood	-1706.844	-1705.130	-1704.735	-1703.165	-1706.057	-1704.110	-1,703.278	-1,702.127

Table reports weighted marginal effects; associated robust standard errors are in parentheses. *** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percentlevel and * denotes statistical significance at the 10 percent level. The dependent variable is the total employment in an establishment in the last period. Initial period variables include spillovers, unemployment rate, and county high-tech wages. See text for details.

Table 6: Censored log–linear regression results for software employment growth of incumbents at a location

Variable	Log(*final employment*)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log initial software employment:	0.528***	0.540***	0.479***	0.492***	0.427***	0.423***
within a mile	(0.066)	(0.066)	(0.074)	(0.073)	(0.088)	(0.088)
Log initial software employment:		0.017	-0.029	0.092	0.021	0.094
1 - 5 miles		(0.051)	(0.054)	(0.090)	(0.121)	(0.140)
Log initial software employment:		-0.152*	-0.164**	-0.200	-0.368*	-0.329
5 - 10 miles		(0.084)	(0.082)	(0.160)	(0.207)	(0.220)
Log initial software employment:		0.107	0.092	0.186	-0.101	0.041
10 - 25 miles		(0.072)	(0.071)	(0.152)	(0.121)	(0.160)
Log initial weighted other industry			-0.099	-0.109	-0.081	-0.046
employment: within a mile			(0.170)	(0.179)	(0.180)	(0.184)
Log initial number of software					0.745**	0.888***
establishments: within a mile					(0.339)	(0.336)
Log initial number of software					-0.185	0.366
establishments: 1 - 5 miles					(0.291)	(0.506)
Log initial number of software					0.333	0.647
establishments: 5 - 10 miles					(0.380)	(0.504)
Log initial number of software					0.371	0.604
establishments: 10 - 25 miles					(0.294)	(0.478)
Log final number of other software						-0.402
establishments: 1 - 5 miles						(0.415)
Log final number of other software						-0.554
establishments: 5 - 10 miles						(0.581)
Log final number of other software						0.217
establishments: $10 - 25$ miles						(0.560)
Log final software employment:				-0.163*		-0.167
$1 - 5$ miles				(0.085)		(0.141)
Log final software employment:				-0.001		0.016
$5 - 10$ miles				(0.144)		(0.195)
Log final software employment:				-0.143		-0.309
$10 - 25$ miles				(0.150)		(0.208)
Log final weighted number of other			0.238	0.284	0.170	0.174
industry employment: within a mile			(0.183)	(0.194)	(0.201)	(0.197)
Initial period variables		Yes**	Yes**	Yes**	Yes**	Yes**
Final period variables				Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Number of obs.	201	201	201	201	201	201
Number of uncensored obs.	136	136	136	136	136	136
Log likelihood	-330.735	-326.026	-324.005	-318.980	-316.960	-312.921

Table reports marginal effects; associated robust standard errors are in parentheses. *** denotes statistical significance at the 1 percent level, ** denotes statistical significance at the 5 percentlevel and * denotes statistical significance at the 10 percent level. The dependent variable is the log of total employment in non-overlapping locations in the last period. Initial and final period variables include spillovers, unemployment rate, and county high-tech wages. Final variables are lagged by one period. See text for details.

Table 7: Censored linear regression results for software employment growth of incumbents at a location

Variable	Final employment					
	(1)	(2)	(3)	(4)	(5)	(6)
Initial number of software	0.539**	0.541**	0.502**	0.499**	0.444**	0.431**
employment: within a mile	(0.240)	(0.229)	(0.214)	(0.208)	(0.187)	(0.183)
Initial number of software		0.004	-0.005	0.103*	-0.045	0.088
employment: 1 - 5 miles		(0.019)	(0.018)	(0.061)	(0.040)	(0.055)
Initial number of software		-0.027	-0.025	0.048	-0.034	0.062
employment: 5 - 10 miles		(0.021)	(0.021)	(0.034)	(0.030)	(0.042)
Initial number of software		0.032	0.034	0.085	0.021	0.081**
employment: 10 - 25 miles		(0.027)	(0.028)	(0.054)	(0.016)	(0.040)
Initial weighted number of other			-0.097*	-0.087	-0.078	-0.045
industry employment: within a mile			(0.057)	(0.058)	(0.055)	(0.060)
Initial number of software					40.267**	48.936**
establishments: within a mile					(17.128)	(19.702)
Initial number of software					2.401	8.239
establishments: 1 - 5 miles					(2.319)	(7.534)
Initial number of software					-0.753	5.671
establishments: 5 - 10 miles					(1.657)	(4.056)
Initial number of software					1.533	0.319
establishments: 10 - 25 miles					(2.137)	(3.800)
Final number of other software						-1.770
establishments: 1 - 5 miles						(5.853)
Final number of other software						-5.846
establishments: 5 - 10 miles						(4.612)
Final number of other software						4.272
esablishments: $10 - 25$ miles						(4.766)
Final software employment:				-0.103*		-0.175*
$1 - 5$ miles				(0.059)		(0.094)
Final software employment:				-0.066*		-0.090
$5 - 10$ miles				(0.038)		(0.058)
Final software employment:				-0.046		-0.084
$10 - 25$ miles				(0.032)		(0.052)
Weighted number of other industry			0.089**	0.089***	0.041	0.018
employment: within a mile			(0.037)	(0.034)	(0.035)	(0.045)
Initial period variables		Yes**	Yes**	Yes**	Yes**	Yes**
Final period variables				Yes**	Yes**	Yes**
Major county effects	Yes**	Yes**	Yes**	Yes**	Yes**	Yes**
Number of obs.	201	201	201	201	201	201
Number of uncensored obs.	136	136	136	136	136	136
Log likelihood	-994.816	-987.977	-986.327	-981.557	-981.465	-970.917

Table reports weighted marginal effects; associated robust standard errors are in parentheses.
*** 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 total employment in non-overlapping locations in the last period. Initial and final period variables include spillovers, unemployment rate, and county high-tech wages. Final variables are lagged by one period. See text for details.

Table 8: Ordered probit regression results for software establishment entry

Variable	Number of new software entrants						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of lagged number of software establishments: within a mile	0.372***	0.328***	0.352***				0.264***
	(0.047)	(0.049)	(0.048)				(0.088)
Log of lagged number of software establishments: $1-5$ miles	0.205***	0.192***	0.178***				0.015
	(0.042)	(0.043)	(0.042)				(0.071)
Log of lagged number of software establishments: $5-10$ miles	0.041	0.040	0.020				0.217**
	(0.048)	(0.049)	(0.048)				(0.097)
Log of lagged number of software establishment: $10-25$ miles	0.083**	0.095***	0.010				-0.105
	(0.034)	(0.035)	(0.035)				(0.072)
Log of lagged software employment: within a mile				0.119***	0.104***	0.111***	0.035
				(0.015)	(0.015)	(0.015)	(0.029)
Log of lagged software employment: $1-5$ miles				0.111***	0.100***	0.100***	0.077***
				(0.016)	(0.017)	(0.016)	(0.029)
Log of lagged software employment: $5-10$ miles				0.001	0.002	-0.016	-0.099**
				(0.019)	(0.019)	(0.019)	(0.044)
Log of lagged software employment: $10-25$ miles				0.058***	0.060***	0.025	0.060*
				(0.015)	(0.015)	(0.016)	(0.035)
Log of lagged weighted other industry employment: within a mile	0.119***	0.143***	0.127***	0.127***	0.152***	0.134***	0.123***
	(0.017)	(0.018)	(0.017)	(0.016)	(0.017)	(0.017)	(0.018)
Log of lagged university spillover			0.021			0.023*	0.019
			(0.013)			(0.013)	(0.013)
Log of lagged junior college spillover			-0.027*			-0.027*	-0.024
			(0.016)			(0.016)	(0.016)
Log of lagged average wage of high-tech industries in the county			0.753***			0.774***	0.763***
			(0.127)			(0.124)	(0.128)
Lagged county unemployment rate			-0.078***			-0.073***	-0.076***
			(0.028)			(0.028)	(0.028)
Lagged county amenity LQ			-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. See text for details.
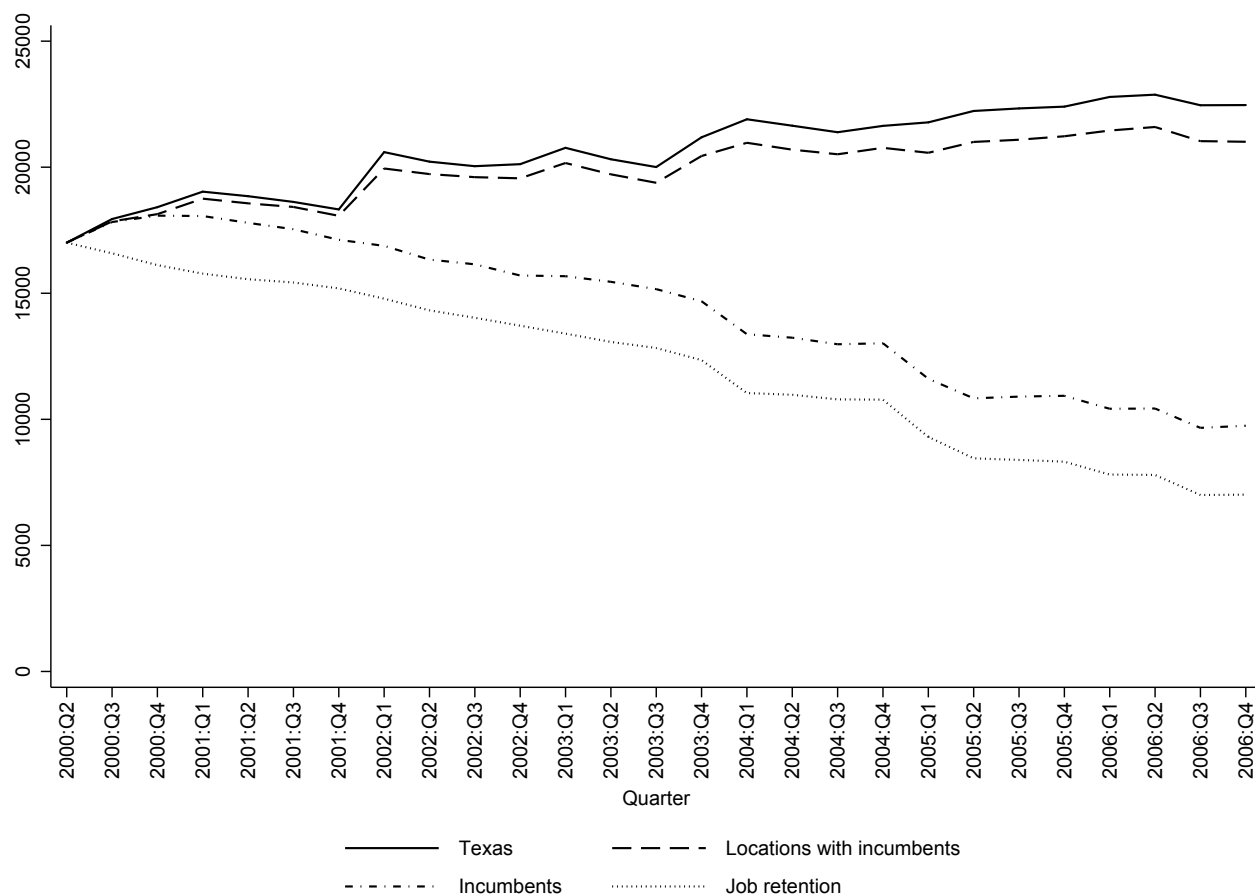
Figure 1. Software Publisher Employment in Texas, 2000-2006. From top-to-bottom: (a) State-wide industry employment; (b) employment in (two mile wide) locations with software publisher presence in 2000-Q2; (c) employment in software publishers active in 2000-Q2; (d) employment in software publishers active in 2000-Q2, excluding employment growth at the establishment level.
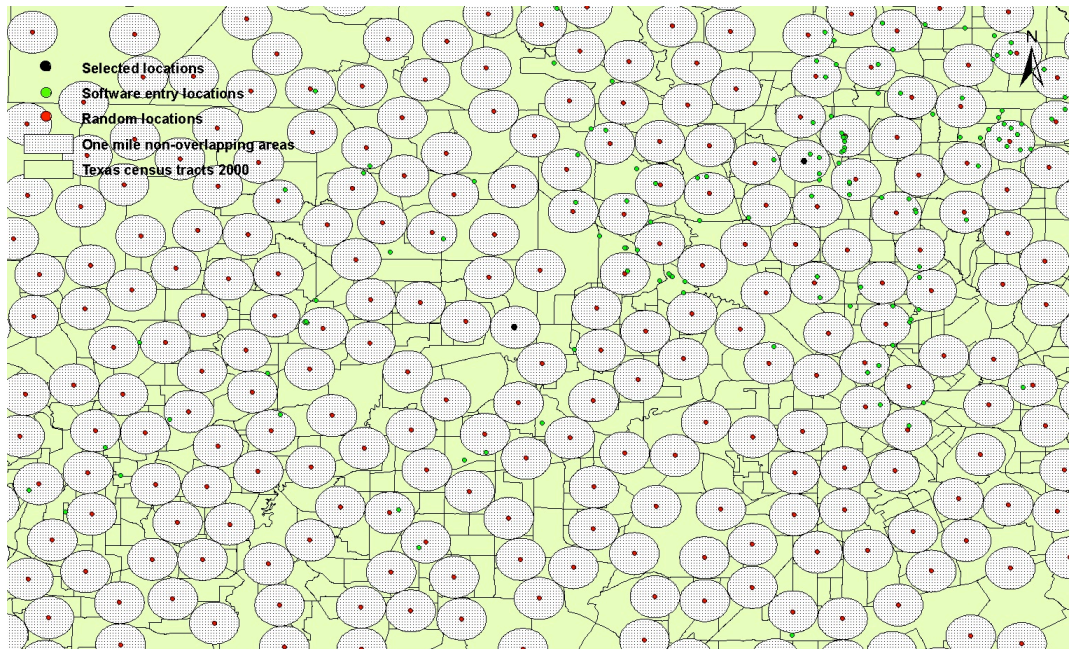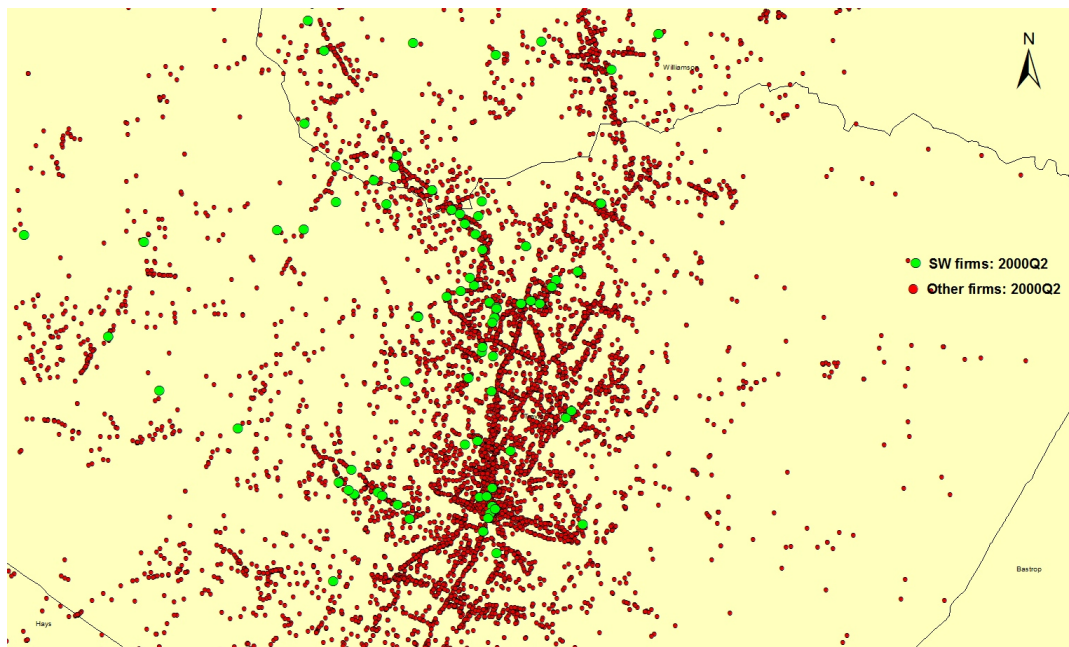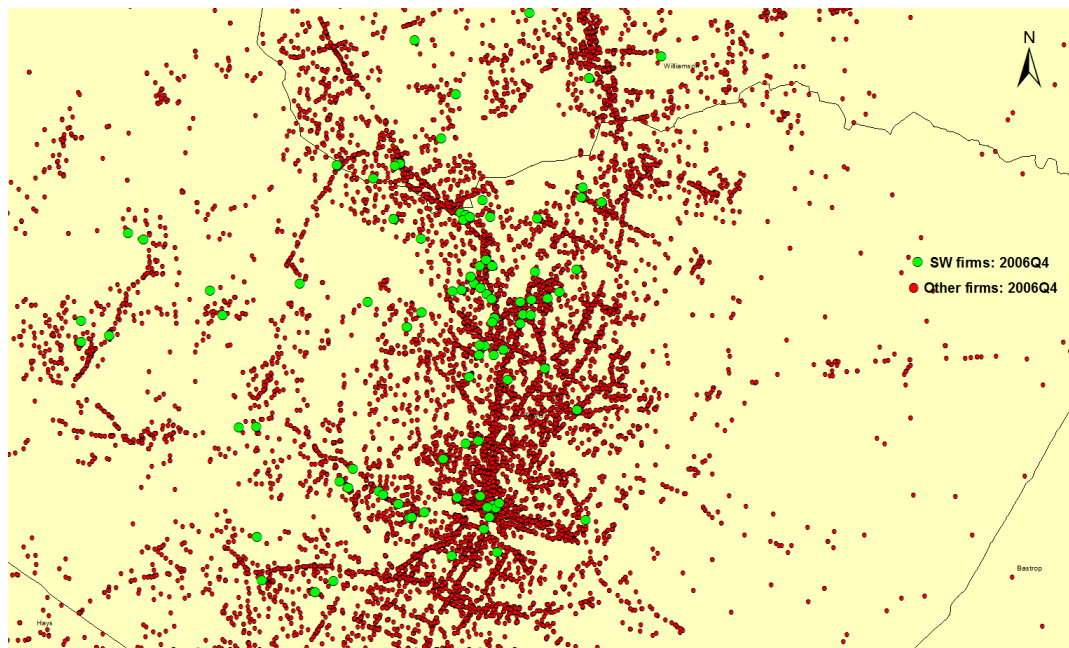
34

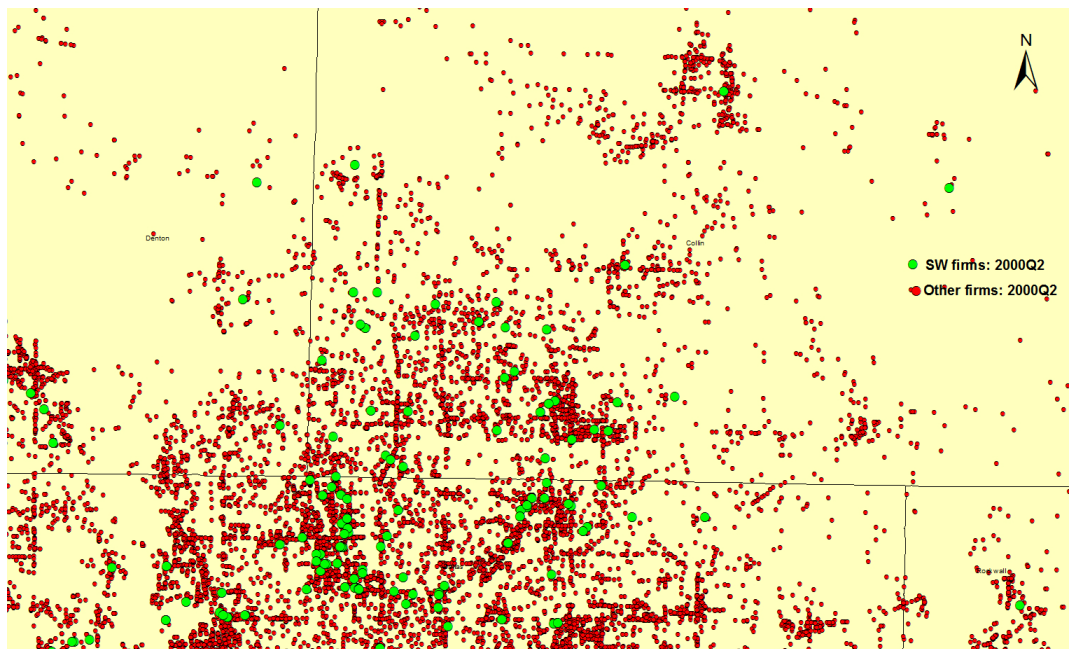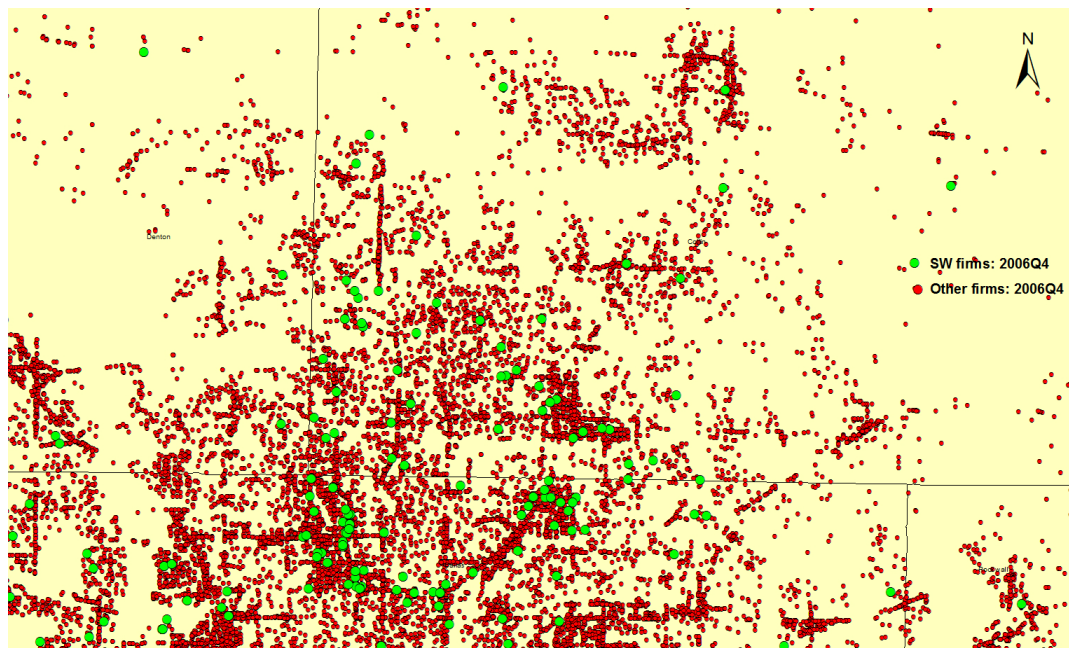Figure 2. Non-overlapping and Selected Locations in Dallas – Fort Worth

Panel A. 2000 Q2

Panel B. 2006 Q4

Figure 3. Distribution of Software and Other Establishments in Austin
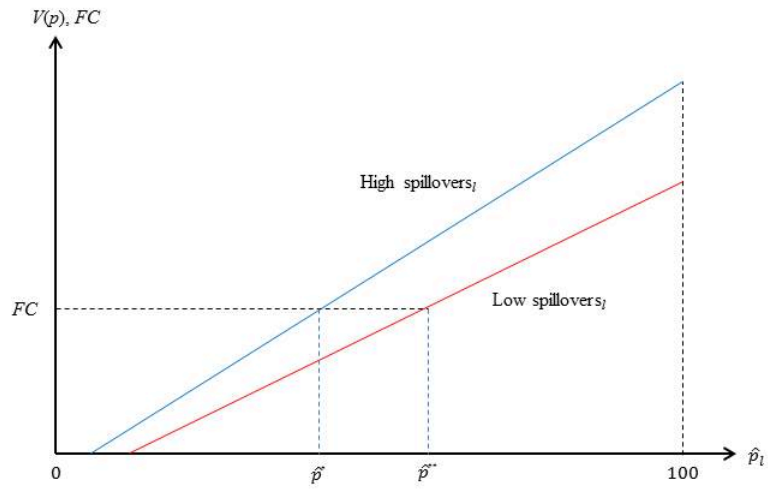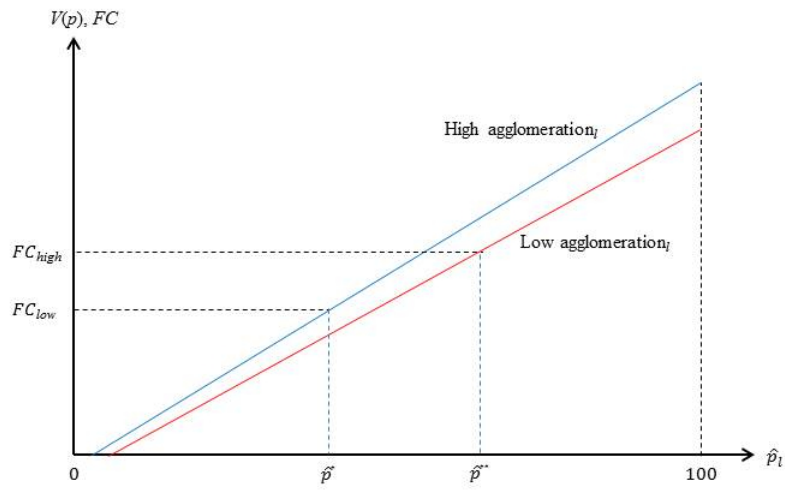
Panel A. 2000 Q2

Panel B. 2006 Q4

Figure 4. Distribution of Software and Other Establishments in North Dallas

Panel A. High and low knowledge spillover locations

Panel B. High and low labor market agglomeration economies

Figure A-1. Profitability and Entry Equilibrium in High and Low Spillover Locations

38