Clustering in the Creative Industries: Insights from the Origins of Computer Software

MARTIN CAMPBELL-KELLY, MARINA DANILEVSKY, DANIEL D. GARCIA-SWARTZ, AND SHANE PEDERSON[1]

University of Warwick and LECG, LLC

Industry and Innovation 17(4), 2010

Abstract

We use several different sources (a 1970 Roster of Organizations in Data Processing and the 1960 and 1970 Censuses of Population) to study patterns of geographic clustering at the very origins of the software industry. We find a strong trend toward clustering of the industry in a few metropolitan areas. Furthermore, we uncover a tendency in the early software industry to agglomerate in close proximity to some of its main customers. This tendency holds even after controlling for region-specific heterogeneity and for the potentially endogenous nature of the software customers' location decisions. We explore the factors that may have driven the observed clustering patterns and suggest directions for further research.

Keywords

Industry studies (services): information and Internet services, computer software

---

[1] Martin Campbell-Kelly is a professor at the Computer Science Department of Warwick University in England. Daniel D. Garcia-Swartz and Shane Pederson are with LECG, LLC. At the time when this study was carried out, Marina Danilevsky was with LECG, LLC as well. Contact Daniel D. Garcia-Swartz, LECG, LLC, 33 West Monroe, Suite 2300, Chicago, IL USA 60603 (email: dgarcia-swartz@lecg.com). We received extremely helpful comments from four anonymous reviewers and from Dr. Mark Lorenzen, the chief editor of this journal. The opinions expressed in this paper are exclusively ours and do not necessarily coincide with those of the institutions with which we are affiliated.

Clustering in the Creative Industries: Insights from the Origins of Computer Software

Abstract

We use several different sources (a 1970 Roster of Organizations in Data Processing and the 1960 and 1970 Censuses of Population) to study patterns of geographic clustering at the very origins of the software industry. We find a strong trend toward clustering of the industry in a few metropolitan areas. Furthermore, we uncover a tendency in the early software industry to agglomerate in close proximity to some of its main customers. This tendency holds even after controlling for region-specific heterogeneity and for the potentially endogenous nature of the software customers' location decisions. We explore the factors that may have driven the observed clustering patterns and suggest directions for further research.

1 Introduction and summary

This paper uses the origins of the software industry to examine an important debate in the regional-economics literature on the drivers of industrial agglomeration. More specifically, our study addresses three questions. In the first place, did the computer software industry in the U.S. exhibit a natural tendency to cluster from the moment of its formation? Secondly, if it did, did it cluster in a specific way? Or put differently, are there any clustering patterns that emerge? And thirdly, if it did cluster in a peculiar way, what were the forces driving the clustering patterns that we observe?

In order to address these questions we resort to two types of data—Census data for 1960 and 1970, and a 1970 Roster of Organizations in Data Processing published in *Computers and Automation*, an industry journal. As regards the first question, we find a strong trend among early software companies (and early software-industry employment) to cluster in a few metropolitan areas. Regarding the second question, we uncover an agglomeration pattern in which early software firms clustered in close proximity to their main customers. Finally, as far as the third question is concerned, toward the end of this study we explore the factors that could plausibly account for this phenomenon.

Our study is organized as follows. Section 2 presents a brief review of the relevant literature, and section 3 focuses on the historical origins of the software industry. In section 4 we describe our data sources, and in section 5 we take a first look at the clustering of the software industry on the basis of a 1970 roster of data-processing organizations. Section 6 takes a deeper look at software-industry clustering in light of data from the 1970 Census. In section 7 we identify the heavy demanders of software products and services in the early decades of the industry, and explore econometrically the co-location of the software industry with some of its main customers. Section 8 focuses on the analysis of potential estimation issues, and section 9 summarizes our conclusions and lays out directions for further research.

2 Literature review

Many of the ideas in the clustering literature have their roots in Marshall (1920), who identified three specific sources of agglomeration economies (i.e., forces driving industrial cluster formation based on increasing returns to scale in the long run): input sharing, knowledge spillovers, and labour market pooling.[2] All of these forces are assumed to increase the productivity of firms that locate in close proximity to other firms belonging to the same industry (see Rosenthal and Strange, 2003; and McCann, 2001: chapter 2). Input sharing focuses on the idea that downstream firms in an industry that is geographically concentrated have the possibility to purchase inputs from specialized suppliers who, in turn, manage to achieve an efficient scale of production precisely due to the agglomeration of downstream companies. Knowledge spillovers refer to the informal exchange of knowledge that takes place among firms that are geographically agglomerated. Labour market pooling can be interpreted as suggesting that workers are better matched to jobs (or companies) in geographic locations that exhibit industrial agglomeration.

The Marshall-oriented line of research focuses on localized externalities. More specifically, it relies on the assumption that, due to a number of factors that are external to any given company, firms prefer to locate in close proximity to other firms in their own industry or

[2] Rosenthal and Strange (2003) provide a comprehensive survey of econometric approaches and findings in this literature, whereas Duranton and Puga (2003) cover some of the main theoretical issues.

in closely related industries. A very intuitive explanation of the logic of Marshall-styled agglomeration economies can be found in Krugman (1991: chapter 2), although Krugman himself is not (or was not) particularly optimistic about finding evidence of knowledge spillovers.

There is, however, a strand of literature that offers a somewhat different interpretation of the forces driving industrial clustering. Krugman (1991: chapter 1), for example, argues that industrial agglomeration arises because of the existence of demand linkages among firms belonging to different industries. The Krugman (1991: chapter 1) model is one where scale economies are internal to the firm, and transportation costs play a crucial role in determining the location of firms. In this model each firm produces one good in a monopolistically competitive setting. Because of the existence of high fixed costs of production, firms tend to concentrate production in one location only; and because of the existence of high transportation costs, each firm locates in close proximity to a large market for its product.

Hanson (2000: 7) has noted that both strands of literature predict industrial clustering. They differ substantially, however, in their interpretation of how clustering happens. In the Marshall-styled models, firms locate in close proximity to other firms belonging to the same industry, whereas in the Krugman-styled models, firms locate close to large consumer and industrial markets and, more specifically, close to large markets for their own products.

Furthermore, the “creative class” strand and the “entrepreneurship” strand offer somewhat different perspectives on clustering. Some of the main tenets and findings of the “creative class” strand are summarized in Florida (2002, 2008). Recent studies in this line of research include Stolarick and Florida (2006), Knudsen, Florida, Stolarick, and Gates (2008), and Florida, Stolarick, and Mellander (2008). Without denying that a number of Marshall- and Krugman-styled forces are often at work, this strand emphasizes the fact that the research focus should be not so much on companies but rather on people. In other words, it is not that creative people cluster because the companies in which they work cluster, but the other way around—creative companies tend to form (and cluster) where creative people tend to agglomerate. In addition, creative individuals tend to cluster in regions that are endowed with certain characteristics.

The "entrepreneurship" strand also suggests that the research focus should shift from companies to individuals, and more specifically to knowledge workers (like scientists and engineers) that, under certain conditions, make the decision to leave their old companies and start new ones (Audretsch, 1995; Audretsch and Feldman, 2004). From this perspective, entrepreneurial activity is more intense in regions where there are large agglomerations of knowledge workers precisely because these agglomerations facilitate access to tacit knowledge (i.e., knowledge spillovers) and thus increase the expected profits from entrepreneurial activity. In other words, companies (and particularly knowledge-intensive companies) usually appear to be clustered together because it is precisely in knowledge-intensive clusters where new companies tend to be born most often.[3]

3 Historical background on the origins of the software industry

Our focus is neither on whether the software industry tends to cluster in the present day—there is some evidence that it does—nor on whether agglomeration in software increases its productivity. We explore clustering at the origins of the software industry, namely in the 1960s.

Until the mid 1960s the software industry as such was quite limited. Software was produced by computer manufacturers, a few software companies, and the customers themselves—and customers accounted for most of it (Phister, 1979: 277).[4] The history of computing hardware and software changed in 1965 with the introduction of IBM's System/360, a family of machines that were software-compatible throughout the performance range. Software written for any machine type within the family worked well on any other machine type, which eliminated switching costs for customers as long as they stayed within the confines of System/360 (Pugh, 1996: chapters 18 and 19).

---

[3] In recent years this journal has published a number of articles on topics related to our study. See, for example, Visser (2009), Lange et al (2008), Lazzeretti, Boix, and Capone (2008), Staber (2008), Dahlander, Frederiksen, and Rullani (2008), Di Maria and Finotto (2008), Kaiser and Muller-Seitz (2008), and Kesidou, Caniels, and Romijn (2009).

[4] Our study focuses on the origins of the software industry in the U.S. For an international perspective see the studies in D. Mowery (1996).

System/360 also had phenomenal implications for the software industry. The number and variety of companies adopting computers to meet specific business needs grew by leaps and bounds after the introduction of System/360, and with the growth of computer hardware installations came a rising demand for computer software. Between 1965 and 1970 hundreds of companies entered the market to supply software products and services for the new hardware platform (and its competitors). FIGURE 1 is based on the 1970 Roster of Organizations in Data Processing (that we describe in the next section). The figure tracks the number of companies that entered into the production and sale of software products and services between 1950 and 1970. It reflects the explosion in entry into software that took place starting around 1965.[5]

<FIGURE 1 ABOUT HERE>

4 Data sources

In our study of clustering in the early software industry we use two types of sources: the "Roster of Organizations in Data Processing" published in the November 1970 issue of *Computers and Automation* ("the 1970 Roster"), on the one hand, and the 1960 and 1970 Censuses, on the other.

The 1970 Roster was intended to be a full census of organizations engaged in data processing, but ended up becoming a sample since many of the organizations did not send their survey responses. In any case, to our knowledge it is the most comprehensive source of this kind that exists for the origins of the software industry. We rely on it under the assumption that the distribution of the relevant characteristics among the non-respondents is not significantly different from that of the respondents.

[5] The acceleration of entry into software is, we believe, clearly associated with the advent of System/360. Among the software companies in the 1970 Roster that provide information on the computing hardware for which they produce/sell software, IBM System/360 is the hardware of choice for 57 percent of companies (including those that report IBM System/360 as well as another hardware system).

The 1970 Roster classifies companies on the basis of a number of categories: computing and data processing services, time-shared computing services, instruction in computer programming, consulting services in the computer field, leasing of computing and data-processing equipment, and selling or producing software. A company could belong simultaneously to several of these categories, and most of them did. In fact, most of the companies that "sold or produced software" were also computer services companies. The 1970 Roster provides information on each company, its 1970 address (city, state, zip code), the year in which it was founded, the number of employees it had in 1970, and a brief description of its activities.

We also relied on the 5-percent samples from the 1960 and 1970 Census of Population and Housing (Ruggles et al, 2008). The industrial classification codes for computer hardware and software show up for the first time in the 1970 Census, but we also used the 1960 Census in some of the econometric analyses. We processed the electronic Census files that are available on the IPUMS website. In order to identify the software industry and its main customers we used the IND1990 variable.[6]

[6] The IND 1990 codes are available at http://usa.ipums.org/usa-action/codes.do?mnemonic=IND1990, last visited on July 30, 2009. We used code # 732 for the software industry, code # 292 for ordnance, code # 322 for the manufacturers of computer hardware, code # 352 for aircraft and parts, code # 700 for banking, code # 711 for insurance, code # 890 for accounting, code # 892 for management, and code # 932 for national security. One caveat is worth mentioning: code # 732 includes companies that supplied computer software and computer services. The 1970 Roster reveals that most companies that supplied computer software also delivered some sort of computer service, and most companies that supplied computer services were also involved in the production and/or sale of software. Apparently there were, however, more companies supplying services but not software than there were companies supplying software but not services. Thus, we raised the question: could we be introducing biases in our calculations because of the existence of some companies that delivered computer services but not computer software? The 1970 Census does not allow us to explore this issue, but the 1970 Roster does. We sorted the metropolitan areas in two ways: first, we ordered them by the total number of people employed in companies that supplied services generally (including those employed in companies that supplied both services and software); then, we ordered them by the total number of people employed in companies that supplied software (regardless of whether they also supplied a computer service or not). The correlation coefficient between these two orderings was about 0.99. In short, if there are any biases in our calculations arising from our use of code # 732 for classifying software-industry companies, they are trivial.

5 A first look at software clustering

Did the early computer companies tend to cluster in specific locations? Here we take a first look at the issue on the basis of the 1970 Roster.

In this first pass we ask two questions, the first one focused on the number of companies or organizations and the second one focused on the number of software workers. First, did companies tend to agglomerate in specific regions? Secondly, did employment in the software industry tend to cluster in specific regions?

TABLE 1 provides information on the top 20 regions (3-digit zip codes) in terms of the number of software companies located in each area in 1970 (and the percent of all software companies accounted for by the region). The 100 3-digit area code group, in New York City, accounted for roughly 11 percent of the companies, and the 606 zip code group, in Chicago, for an additional 4 percent or so. Each one of the next three regions—Cambridge, MA, Washington, DC, and Los Angeles, CA—accounted for between 3 and 4 percent of all software companies. The vast majority of the software companies located in these areas came to life in 1965 or later, and were not very large. The top 10 areas accounted for more than 34 percent of all companies, and the top 20 for more than 46 percent, which suggests that geographic clustering constituted a defining feature of the early software industry.

<TABLE 1 ABOUT HERE>

TABLE 2 provides information on the top 20 regions (3-digit zip codes) in terms of the number of software-industry employees working in the region in 1970 (and the percent of all software-industry employees accounted for by the region). The table indicates that roughly 20 percent of the employees worked in Los Angeles, CA. The number two spot belonged to New York, NY, with 14 percent of the software work force, but the next three spots were once again in California. This time a few large firms drive a good portion of the results. For example, in terms of number of employees the top two firms in the 1970 Roster were in Los Angeles. They had many more employees than the largest firm in Chicago, IL, which only had 100 in 1970. The top 10 areas accounted for 63 percent of all software-industry employment, and the top 20 for more than 75 percent. Clustering in software in terms of industry employment was even starker than in terms of the number of firms.

<TABLE 2 ABOUT HERE>

6 A deeper look at software clustering

Having uncovered strong evidence of clustering in the 1970 Roster, we here make use of Census information to study the phenomenon from a somewhat different perspective. More specifically, the 1970 Census of Population and Housing (Ruggles et al, 2008) allows us to study clustering in the software industry not through a sample of establishments but through a sample of households. In addition, it allows us to more carefully explore some of the driving forces behind early clustering in the software industry. We use the 1970 Census because it captures the geographic localization of the software industry at the end of the explosive bout of entry that took place between 1965 and 1970.

6.1 Location quotients

For each sampled individual, the 1970 Census records the geographic area in which she lived and, if she was employed, the industry in which she worked. Thus, it allows us to calculate location quotients for software-industry employment. The location quotient for an industry attempts to measure the geographic concentration of that industry. Recall that the location quotient for industry j in region r is $LQ_{j,r}=\left(\frac{E_{j,r}}{E_r}\right)/\left(\frac{E_j}{E}\right)$, where $E_{j,r}$ is the total number of individuals employed in industry j and region r, E_r is total employment in region r (regardless of industry), E_j is total employment in industry j (regardless of region), and E is total economy-wide employment (regardless of industry and region) (McCann, 2001: 144-146).

The economic logic behind location quotients is that if a region has an employment share in any given industry that is greater than the national average, then the region must be specialized in relative terms in the production of the output of that industry (McCann, 2001: 144). A location quotient equal to 1 for an industry in a region indicates industry employment at the "expected" level in that region relative to total employment. Some authors consider a location quotient greater than 1 to be a sign of industry agglomeration in a region; others prefer to view location quotient measures as a continuum—the higher the location quotient for an industry in a region, the more intense the agglomeration of that industry in that region.

6.2 Findings

On the basis of the 1970 Census we explore the presence of clustering in the software industry across metropolitan areas. Among the 128 metropolitan areas in the Census, 40 have a location quotient for the software industry greater than unity. The mean location quotient for the software industry is 0.907, the median is 0.686, the minimum is zero, and the maximum is 9.216. TABLE 3 presents the top 20 metropolitan regions in terms of the location quotient for the software industry with the total employment of each region (in all industries).

<TABLE 3 ABOUT HERE>

In general, the 1970 Census and the 1970 Roster agree in terms of the areas where the software industry was heavily clustered at the end of the 1965-1970 expansion. Metropolitan areas like New York, Chicago, Los Angeles, Cambridge-Boston, Washington DC, Philadelphia, Minneapolis, Dallas, Houston, and Palo Alto/San Jose appear consistently at the top, no matter how we look at software clustering.[7]

7 Was there a pattern to software-industry clustering?

Is there a discernible pattern to the way in which the software industry clustered in the early years of its existence? We explore the idea that the software industry clustered in close proximity to its main customers. Note that this is theoretically consistent with Krugman (1991: chapter 1), but it is not inconsistent with elements of the other approaches that we surveyed in the literature-review section.

7.1 Identifying the heavy demanders of software products and services

In order to study whether the software industry clustered in close proximity to its main customers, we need first to identify which industries demanded heavily software products and services in the 1960s. In order to accomplish this we can proceed in a number of ways. First, we can look at sources that provide information on the sectors of the economy that were heavy demanders of computer hardware. Since the services of computer hardware and software are complements relative to each other, heavy demanders of hardware services were also likely heavy demanders of software services.

In this regard there are at least two approaches. First, we can look at those sectors of the economy that relied heavily on computers in the sense that a large proportion of all establishments in the sector had computer installations. From this perspective, certain manufacturing sectors (particularly durable good manufacturers), and banking and financial services appear at the top of the list (Phister, 1979: 447).

Secondly, we can look at those sectors of the economy that accounted for large portions of the whole stock of computer installations in the country. From this perspective, certain discrete manufacturing sectors, certain business services, finance, insurance, real estate, and the government sector show up at the top of the list (Phister, 1979: 444-445).

There are also alternative approaches based on Census data. We can ask the question: Outside of the computer software industry, what other sectors of the economy demanded heavily the services of computer programmers? The sectors that heavily employed programmers in-house were likely also sectors that purchased software services from software companies when the in-house programming staff did not have the skill set that was required for specific tasks. From this perspective, we can ask two different questions. The first one is: What sectors of the economy (other than the computer software industry) employed large portions of all computer programmers in the economy? The second one is: What sectors of the economy (other than the computer software industry) exhibited high "industrial" location quotients for software programmers? Answering the second question requires that we calculate the location quotient for computer programmers in each sector of the economy (other than the software industry).

We carried out calculations based on the 1970 Census data following the two approaches. There were a number of sectors (8, more specifically) that showed up at the top of the lists

[7] Some of the discrepancies may arise from two sources. For one, the 1970 Roster is a sample of establishments whereas the 1970 Census is a sample of households. Furthermore, location quotients involve scaling, in such a way that the size of the region (in terms of employment) is taken into account.

constructed with both. They were (1) national security, (2) the manufacturing of computer equipment, (3) banking, (4) insurance, (5) accounting, auditing, and bookkeeping, (6) management and public relations, (7) ordnance, and (8) aircraft and parts.

In very general terms these sectors tend to coincide with the ones identified previously on the basis of the demand for hardware services. Furthermore, the information we have on the evolution of the early computer industry suggests that these were the sectors that computerized most heavily early on. Banks and insurance companies were among the top IBM customers in the early decades (Yates, 2005; Cortada, 2006). The manufacturers of computer equipment themselves often contracted with the early software companies for the creation of specific pieces of software (Campbell-Kelly, 2003). Ordnance and aircraft and parts were fundamental portions of a high-technology defence sector that used computers intensely from the beginning. Computers were heavily used for a wide variety of tasks related to national security in the 1950s and 1960s (Flamm, 1988).

Accounting, auditing, and bookkeeping, on the one hand, and management and public relations, on the other, are two "sectors" that raise some questions. The evidence does suggest that these sectors computerized early on, and thus may have been heavy demanders of software in the 1960s. There is an alternative explanation, however, for the potential co-location of software-industry employment with accounting and management employment—all of these sectors were important business-service sectors, and they could all be locating in close proximity to, say, banking and insurance companies. Thus, on the basis of Census data we conduct our analyses both with the 8 heavy demanders and with 6 of the heavy demanders only (excluding accounting and management).

7.2 Testing hypotheses regarding the co-location of software companies with the heavy demanders of software services

Once we have calculated geographic location quotients for the software industry and for the heavy demanders of software services, we can proceed to test hypotheses regarding a possible pattern of co-location. As a first pass, we computed the correlation coefficient between the location quotient of the software industry and the location quotient of each one of the heavy demanders. We found that the location quotient of the computer software industry is positively correlated with the location quotient of all the heavy demanders of software services except for aircraft and parts, although the correlation coefficients for insurance and ordnance are not statistically significant, and the one for banking is significant at about 7 percent.[8] This evidence, albeit quite preliminary indeed, is consistent with the fact that at the origins of the software industry there is a pattern of co-location of the software industry with the sectors that heavily demanded software products and services.

7.2.1 Model # 1

In order to explore this evidence further, we used three different regression models, two of which are, in our view, somewhat unsatisfactory for different reasons. Model # 1 is a regression of the location quotient for the software industry on all the location quotients for the heavy demanders (in one equation):

$$LQ_r^{SC} = \alpha + \sum_{h=1}^{H} \beta_h LQ_{h,r}^{HD} + \delta EMP_r + \gamma EMP^2 + \varepsilon_r \qquad (1)$$

In this model, LQ_r^{SC} is the location quotient for the software industry in region r, $LQ_{h,r}^{HD}$ is the location quotient for heavy demander h in region r, and $h = 1,\ldots,H$ indexes the sectors that we have identified as heavy demanders of software products and services. The regression also controls for a measure of the region's employment (divided by 10,000) and employment squared (to account for the fact that the impact of regional employment on the location of the software industry may be non-linear). The results from this model are in TABLE 4.[9]

<TABLE 4 ABOUT HERE>

Both regional employment coefficients are statistically significant. The coefficient on the linear term is positive and the one on the squared terms is negative, which suggests that up to a point the software industry tended to locate more intensely in more heavily populated areas. As far as the heavy demanders are concerned, all the coefficients but one (aircraft and parts)

---

[8] The results, not reported here for space reasons, are available from the authors upon request.

[9] All the regressions reported in this study were estimated with robust standard errors. For every model we estimated via OLS we also estimated a TOBIT version to account for censoring in the left-hand-side variable (the location quotient for the software industry). The TOBIT results, however, were not substantially different from the OLS numbers reported in the relevant tables or text. In addition we explored different functional-form specifications. For example, we included the location quotient squared for each one of the heavy demanders on the right-hand-side, but this term turned out to be not statistically significant in most cases. Furthermore, we tried a log-log specification (that is, we estimated a model for the log of the location quotient for the software industry on the log of the location quotient for the heavy demanders), but the results were not substantially different from those reported in the core of this study.

are positive, although only the coefficients on management and national security are significant at standard levels. The problem with a model like this one is that there are patterns of strong correlation among the location quotients of the heavy demanders with one another. For example, management tended to co-locate with accounting and national security; banking tended to co-locate with insurance; and ordnance tended to co-locate with computer manufacturing.[10] Thus, it is far from clear that the coefficients reported in TABLE 4 are informative if we care about the potential co-location of the software industry with the heavy demanders.

7.2.2 Model # 2

Model # 2 offers an alternative approach:

$$LQ_r^{SC} = \alpha + \beta LQ_{h,r}^{HD} + \delta EMP_r + \gamma EMP^2 + \varepsilon_r \qquad (2) \qquad for\ h = 1,\ldots,H$$

In other words, we estimated individual OLS regressions of the location quotient for the software industry on the location quotients for each one of the heavy demanders of software services and the region's total employment. TABLE 5a and TABLE 5b present the results of this exercise.

<TABLE 5a ABOUT HERE>

<TABLE 5b ABOUT HERE>

Again, the coefficients on regional employment are statistically significant. All of the individual coefficients on the heavy demanders but one (aircraft and parts) are positive, and a couple of them (management and national security) are statistically significant at standard levels. A few others are close to being significant. This model is better than Model # 1 inasmuch as it avoids the problems arising from the co-location of the heavy demanders with each other, but it still has a fundamental shortcoming.

Think of a world where there are two industries that heavily demand the services of the software industry, say national security and banking. Assume that their geographic location patterns are negatively correlated (as in fact they are for these two sectors in 1970).[11] Assume also that there are *R* regions in the country, half of them with a heavy concentration of

[10] These results are available from the authors upon request.
[11] These findings are available from the authors upon request.

banking-industry employment (the "banking regions") and the other half of national-security employment (the "security regions"). Now, assume that software companies tend to specialize and a large number of them agglomerate in each region to meet the demands of the heavy demander located there (either banking or national security). Call LQ_r^{SC} the location quotient for the software industry in region r, $for\ r = 1,...,R$, LQ_r^{BK} the location quotient for the banking industry in region r, and LQ_r^{NS} the location quotient for the national security sector in region r. It could happen that neither an individual regression of LQ_r^{SC} on LQ_r^{BK} nor an individual regression of LQ_r^{SC} on LQ_r^{NS} shows a co-location effect, even though the software industry is in fact co-locating with the heavy demanders of its services.

7.2.3 Model # 3

We need a model that allows us to capture the co-location patterns, if they exist, of the software industry with all the heavy demanders of software services simultaneously. In this regard, our preferred model is Model # 3. This approach relies on estimating a location quotient for all heavy demanders together. In other words, we create a location quotient

$LQ_{ALL,r}^{HD} = \left(\frac{\sum_{h=1}^{H} E_{h,r}}{E_r} \right) / \left(\frac{\sum_{h=1}^{H} E_h}{E} \right)$, where $h = 1,...,H$ indexes the sectors that we have identified as heavy demanders of software products and services. Then we estimate a regression model of the location quotient for the software industry on the location quotient for the "sector of high demanders" and the region's employment:

$$LQ_r^{SC} = \alpha + \beta\, LQ_{ALL,r}^{HD} + \delta\, EMP_r + \gamma EMP^2 + \varepsilon_r \qquad (3)$$

Table 6 reports the results from estimating this model, which is our preferred model. The effects are both substantial and statistically significant. A one-unit increase in the location quotient for the combined heavy-demander sector added between 0.624 units (with the 8 heavy demanders combined into one) and 0.544 units (after excluding accounting and management) to the location quotient of the software industry. This is a massive effect indeed, and suggests that there was a strong tendency in the early software industry to cluster geographically in close proximity to some of its main customers.[12]

<TABLE 6 ABOUT HERE>

Model # 3 is consistent with a couple of different stories. One of them is that early software companies tended to specialize in providing services and products to specific sectors of the economy. The higher the number of high-demander sectors that were clustered in a specific region, the more software companies tended to locate there since different subsets of software companies served the needs of different high demanders. An alternative story is that software companies already in the 1960s were becoming quite diversified. Thus, regions where a wide variety of heavy demanders located tended to attract software companies because they provided a hedge against the risk of sector-specific negative shocks—if, say, the demand for national security declined, software companies could still survive from work they did, say, for the banking industry. Our analysis of the economic history of software in the first couple of decades (and of the 1970 Roster) suggests that both stories contain elements of truth, but the second one is more pertinent—software companies were becoming reasonably diversified already in the 1960s.

8 Potential estimation issues

We have uncovered a strong pattern of co-location of software companies (or software-industry employment more generally) with companies that heavily demanded software products and services. We have interpreted this pattern as driven by the fact that software companies tended to agglomerate in close proximity to their main customers. Are there alternative explanations for the co-location pattern? Potentially, there are. Let us explore them briefly.

We can think of two alternative explanations. First, there could be unobserved, region-specific factors that drove the clustering of both the heavy demanders of software and the software companies. The problem here would be one of potential region-specific unobserved

[12] We also estimated a different model (not reported here for space reasons), which suggests that the clustering of one additional heavy demander in a region added between 0.25 and 0.28 units to the location quotient of the

heterogeneity. Secondly, we might have interpreted the causality in exactly the wrong way: it could be the case that the heavy demanders tended to locate in close proximity to the established software companies (because they heavily used their services) rather than the other way around. The problem here would be that the location patterns of the heavy demanders are endogenous vis-à-vis those of the software industry.[13]

8.1 Region-specific unobserved heterogeneity

We would like to address potential region-specific unobserved heterogeneity through some form of first differencing of the data. It is well known that first-difference, or more generally fixed-effects, models are designed to address these kinds of problems (see, for example, Wooldridge, 2002: chapter 10). We explored this possibility with Census 1960 and Census 1970 data. Unfortunately, the 1960 Census does not contain information on the software industry.

Although we cannot estimate a first-difference model on the basis of the available data, we can attempt to control, albeit imperfectly, for unobserved region-specific heterogeneity by including the location quotient for college/university employment in our regressions. The agglomeration of college/university employment is a reasonably good control for a number of reasons. It could be hypothesized that both the software industry and the heavy demanders tended to cluster in regions where highly educated workers agglomerated because both sectors employed them intensely—in other words, college cities and towns provided a deep pool of workers for both the software industry and the heavy demanders. Alternatively, along the lines of the "creative class" approach (Florida, 2002, 2008), it could be hypothesized that the creative workers employed by both the software industry and the heavy demanders agglomerated in cities with heavy college/university employment precisely because these cities were endowed with the characteristics that the "creative class" tends to value most highly.

---

software industry in that region. These results are available from the authors upon request.

[13] Strictly speaking, in both cases there is an endogeneity problem. In the first case, the location patterns of both the software industry and the heavy demanders would be endogenous vis-à-vis some region-specific factor that is potentially unobserved by the econometrician. In the second case, the location patterns of the heavy demanders would be endogenous vis-à-vis those of the software industry.

We estimated our preferred model, Model # 3, but now controlling for the 1970 location quotient for college/university employment in each region.[14] We estimated the regression on the combined high-demander sector in three versions: the first one includes all 8 sectors, the second one includes only the 6 sectors that are available both in the 1970 Census and the 1960 Census (i.e., the 8 minus computer manufacturing and ordnance), and the final one further excludes accounting and management (so that it keeps only insurance, banking, national security, and aircrafts and parts). Although for space reasons we do not report the full results, we found that the inclusion of college employment on the right hand side did not change our findings in any substantial manner.[15]

8.2 Endogenous co-location of the heavy demanders

An additional problem could be, as we pointed out, that we might have interpreted the causality in exactly the wrong way. It could be that the heavy demanders tended to co-locate with the software companies (because they intensely demanded their services) rather than the other way around. A priori this story does not seem entirely plausible, since it seems to imply that the heavy demanders dramatically changed their location patterns between 1960 and 1970 to follow an industry (software) that hardly existed in 1960. There are, however, historical examples of computer manufacturers moving close to software providers (see, for example, Saxenian, 1996: 157). In any case, the issue can be explored econometrically, since data on the location patterns of the heavy demanders are available (for most of them, at least) in the 1960 Census.

There are two approaches to dealing with the potentially endogenous nature of the location patterns of the heavy demanders. First, we can estimate a regression of the indicators of clustering for software companies in 1970 on the indicators of clustering for the heavy demanders in 1960—the location decisions of the heavy demanders in 1960 were clearly exogenous vis-à-vis the location patterns of the software industry in 1970. Secondly, we can run instrumental-variable regressions of the clustering indicators for the software industry in

[14] Data for college/university employment are available for 1970 but not for 1960.
[15] The full results are available from the authors upon request.

1970 on clustering indicators for the heavy demanders in 1970, using the clustering indicators for the heavy demanders in 1960 as instrumental variables. We attempted both approaches.[16]

We report two sets of results. In order to tackle the potentially endogenous nature of the location of the heavy demanders, we first estimated our preferred model (Model # 3) but using the location quotient of the (combined) heavy-demander sector in 1960 on the right hand side. We controlled for the agglomeration of college/university employment and for the overall level of the region's employment (in all sectors). These results are reported in TABLE 7a.

We also explored the issue in an instrumental-variable framework. Here we used the (combined) location quotient for the heavy demanders in 1960 as an instrument for the location quotient of the (combined) heavy demander sector in 1970 in that region. Once again, we controlled for the agglomeration of college/university employment in a specific region and for the region's total employment. These results are reported in TABLE 7b.

<TABLE 7a ABOUT HERE>

<TABLE 7b ABOUT HERE>

All the estimated coefficients on the combined heavy-demander sector were positive and statistically significant. Adding one unit to the location quotient for the combined high-demander sector in 1960 increased the location quotient of software-industry employment in 1970 by between 0.694 and 0.722 units, a substantial and highly significant effect. Furthermore, adding one unit to the instrumented location quotient for the combined high-demander sector in 1970 increased the location quotient of software-industry employment in 1970 by between 0.789 and 0.837 units, again a substantial and highly statistically significant effect. These results provide further evidence that the early software companies tended to cluster in close proximity to their main customers.

[16] Since in both cases we relied on 1960 data, we were forced to estimate the models on the basis of the geographic units identified in the 1960 Census, which are considerably less than those identified in the 1970 Census. This is reflected in the degrees of freedom reported in the relevant tables. In any case, the fact that the coefficients on the combined heavy-demander sector estimated in this context reveal a substantial and statistically significant effect, even with a considerably smaller number of observations, suggests that the relationships we have uncovered are quite robust.

9 Conclusions and directions for further research

We have addressed the first two questions posed in the introduction. That is, we have addressed the issue whether the early software industry clustered or not, and we have explored the clustering patterns that emerged—namely, agglomeration in close proximity to some of its main customers. We have not yet addressed the third question, the one about the forces driving the kinds of clustering patterns we observe. Here we briefly present some hypotheses in this regard, and suggest directions for further research.

In our view, the most powerful explanation for the clustering patterns we observe bears some connections to the Krugman (1991: chapter 1) model. In this regard we have to pose two questions. For one, did fixed costs matter in the early history of software? And additionally, did transportation costs matter?

First, let us focus broadly on "fixed costs." There is some evidence that it was more efficient for a small software company to have all of its programmers working on one or two large and very powerful mainframes in one location than to have them spread out. A number of studies of the computer industry confirmed the existence of a principle known as Grosch's Law—for any given generation of computers, computer performance was proportional to the square of the system's rental, so by purchasing (or renting) a piece of hardware that was twice as expensive the user obtained roughly four times as much computing capacity (Phister, 1979: 62, and Knight, 1968).[17] Therefore, most early software companies tended to concentrate the production of software in one location.

If software production was to be concentrated in one location, was it more efficient to locate close to large customers? This depends in part on whether "transportation costs" (broadly defined) mattered, and in the early days of the software industry they surely did. In the 1960s software companies did not sell the mass-market software packages that became typical in the 1980s—they sold software products that required a substantial amount of pre- and post-sales support (Campbell-Kelly, 2003: chapter 1). Therefore, being close to the customer was crucial for the success of the early software companies. In the early days of the

[17] The advent of the computer timesharing industry (Campbell-Kelly and Garcia-Swartz, 2008a) surely made it easier to develop software remotely, but there were limits to the kinds of software that could be produced in this way.

software industry customer thickness mattered because transportation costs mattered; and transport costs mattered because at the time they involved technical and marketing staff visiting customers to gather/provide information before a sale and to provide support after a sale. For all practical purposes, the early history of clustering in the software industry tends to lend some support to Krugman-styled models of regional demand linkages.

It should be pointed out, however, that this is not the only story that can explain the co-location patterns we observe. The evidence is also consistent with elements of the "creative class" theory, with a story focused on Marshallian externalities, and with an economic story that emphasizes entrepreneurial activity in the creative industries, the main elements of all of which we have summarized in the literature review. Determining which one of these stories provides the most convincing explanation for the observed agglomeration patterns will require further research.

In this regard, it may be worth exploring how the clustering patterns have changed over time—that may shed some light on the deep forces driving software clustering. For example, what happened to the early software companies once the customers they were serving disappeared, or migrated, or became more dispersed, if this happened? Furthermore, the transition toward an industry based on standardized software packages surely has to have led to a world in which face-to-face contacts with customers mattered less. Of course, in the case of standardized software for personal computers, personal interactions with customers are close to non-existent. This does not mean that we should expect the creators of software for personal computers not to be clustered—it just means that clustering in that sector of the industry should not be driven by proximity to customers (however those customers are defined). In addition, sectors of the software industry where proximity to customers matters more (the corporate software products sector) may exhibit clustering patterns that are different from those where it does not matter at all (the personal computer software product sector).

This line of research could be extended to encompass a very recent wave of software companies that sell software-as-a-service (Campbell-Kelly and Garcia-Swartz, 2008b). A priori it would seem that nothing is more impersonal than delivering software via a monthly subscription over the Internet—not much interaction with customers would seem to be involved other than over the phone. Are these companies clustering? And if they are, what clustering patterns are emerging?

A final, and perhaps more complex, question emerges from the fact that the software industry has had different waves (and that the various waves have had different characteristics): Did the clustering of a previous generation of companies have an impact on the clustering of the ones that followed? In other words, when the software companies of the mid 1970s came to life they found a world in which software companies of the prior waves were already clustered in certain regions. Did this matter for the later waves in terms of where they chose to locate? Did they attempt to exploit Marshallian externalities with their location choices? Or were the various waves of software companies so different from each other that knowledge spillovers across waves turned out to be of limited significance?

10 Bibliography

Audretsch, D. (1995) Innovation and Industry Evolution (Cambridge: The MIT Press).

Audretsch, D. and Feldman, M. (2004) Knowledge Spillovers and the Geography of Innovation, in Henderson, J. and Thisse, J., eds., Handbook of Urban and Regional Economics, Vol. 4 (New York: North Holland).

Campbell-Kelly, M. (2003) From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry (Cambridge: The MIT Press).

Campbell-Kelly, M. and Aspray, W. (2004) Computer: A History of the Information Machine (Boulder: Westview Press).

Campbell-Kelly, M. and Garcia-Swartz, D. (2008a) Economic Perspectives on the History of the Computer Time-Sharing Industry, 1965-1985, IEEE Annals of the History of Computing, 30(1).

Campbell-Kelly, M. and Garcia-Swartz, D. (2008b) The Rise, Fall, and Resurrection of Software as a Service: Historical Perspectives on the Computer Utility and Software for Lease on a Network, in: Aspray, W. and Ceruzzi, P., editors, The Internet and American Business (Cambridge: The MIT Press).

Computers and Automation (1970) Main Roster of Organizations in Computers and Data Processing, Computers and Automation, 19(6).

Computer Sciences Corporation (2003) The CSC Story.

Cortada, J. (2006) The Digital Hand, vol. 2: How Computers Changed the Work of the American Financial, Telecommunications, Media, and Entertainment Industries (Oxford: Oxford University Press).

Dahlander, L, Frederiksen, L., and Rullani, F. (2008) Online Communities and Open Innovation: Governance and Symbolic Value Creation, Industry and Innovation, 15(5), pp. 115-123.

Di Maria, E., and Finotto, V. (2008) Communities of Consumption and Made in Italy, Industry and Innovation, 15(2), pp. 179-197.

Duranton, G., and Puga, D. (2003) Micro-Foundations of Urban Agglomeration Economies, NBER Working Paper 9931.

Flamm, K. (1988) Creating the Computer: Government, Industry, and High Technology (Washington DC: Brookings Institution).

Florida, R. (2002) The Rise of the Creative Class (New York: Basic Books).

Florida, R. (2008) Who is your City? (New York: Basic Books).

Florida, R., Stolarick, K., and Mellander, C. (2008) Inside the Black Box of Regional Development – Human capital, the creative class, and tolerance, Journal of Economic Geography, 8(5), pp. 615-649.

Hanson, G. (2000) Scale Economies and the Geographic Concentration of Industry, NBER Working Paper 8013.

Kaiser, S., and Muller-Seitz, G. (2008) Leveraging Lead User Knowledge in Software Development—The Case of Weblog Technology, Industry and Innovation, 15(2), pp. 199-221.

Kesidou, E., Caniels, M., and Romijn, H. (2009) Local Knowledge Spillovers and Development: An Exploration of the Software Cluster in Uruguay, Industry and Innovation, 16(2), pp. 247-272.

Knight, K. (1968) Evolving Computer Performance, 1963-1967, DATAMATION (January), pp. 31-35.

Knudsen, B., Florida, R., Stolarick, K., and Gates, G. (2008) Density and Creativity in U.S. Regions, Annals of the Association of American Geographers 98(2), pp. 461-478.

Krugman, P. (1991) Geography and Trade (Cambridge: The MIT Press).

Lange, B., Kalandide, A., St. Ber, B., and Mieg, H. (2008) Berlin's Creative Industries: Governing Creativity?, Industry and Innovation, 15(5), pp. 531-549.

Lazzeretti, L., Boix, R., and Capone, F. (2008) Do Creative Industries Cluster? Mapping Creative Local Production Systems in Italy and Spain, Industry and Innovation, 15(5), pp. 549-567.

Marshall, A. (1920) Principles of Economics (London: MacMillan).

McCann, P. (2001) Urban and Regional Economics (Oxford: Oxford University Press).

Mowery, D. (1996) The International Computer Software Industry (Oxford: Oxford University Press).

Phister, M. (1979) Data Processing: Technology and Economics (Bedford, Mass.: Santa Monica Publishing Company and Digital Press).

Pugh, E. (1996) Building IBM: Shaping an Industry and its Technology (Cambridge: The MIT Press).

Rosenthal, S., and Strange, W. (2003) Evidence on the Nature and Sources of Agglomeration Economies, prepared for the Handbook of Urban and Regional Economics, Volume 4.

Ruggles, S., Sobek, M., Alexander, T., Fitch, C. Goeken, R., Hall, P., King, M. and Ronnander, C. (2008) Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database] (Minneapolis, MN: Minnesota Population Center [producer and distributor]), available at http://usa.ipums.org/usa.

Saxenian, A. (1996) Regional Advantage: Culture and Competition in Silicon Valley and Route 128 (Cambridge: Harvard University Press).

Staber, U. (2008) Network Evolution in Cultural Industries, Industry and Innovation, 15(5), pp. 569-578.

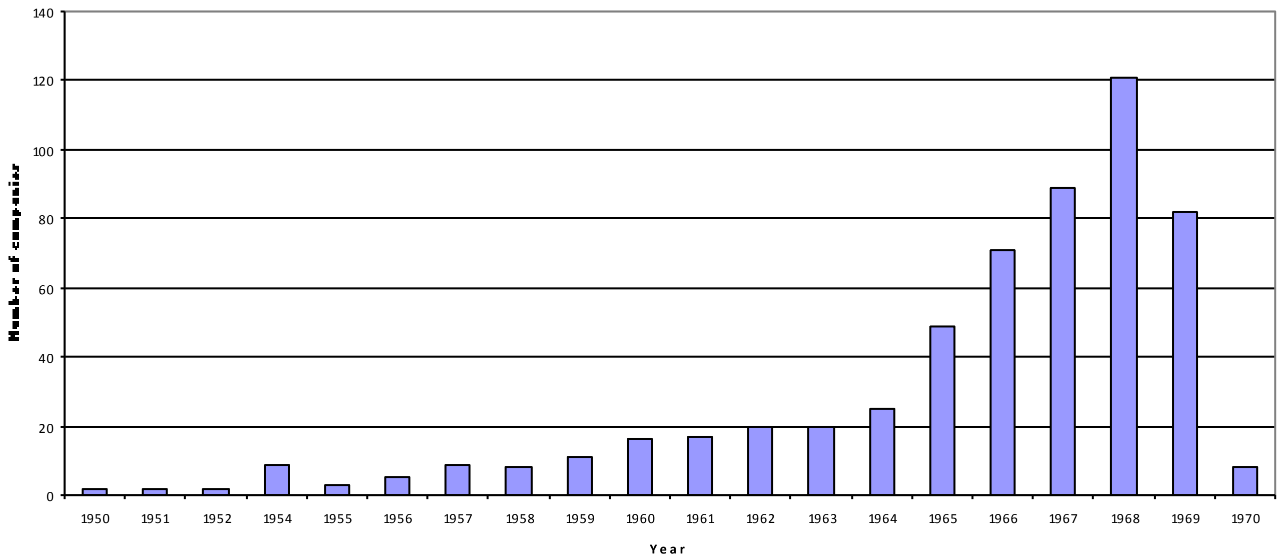
Stolarick, K., and Florida, R. (2006) Creativity, connections and innovation: A study of linkages in the Montreal Region, Environment and Planning A, 38(10), pp. 1799-1817.

Visser, E. (2009) The Complementary Dynamic Effects of Clusters and Networks, Industry and Innovation, 16(2), pp. 167-195.

Wooldridge, J. (2002) Econometric Analysis of Cross Section and Panel Data (Cambridge: The MIT Press).

Yates, J. (2005) Structuring the Information Age: Life Insurance and Technology in the Twentieth Century (Baltimore: Johns Hopkins University Press).

Figure 1. Entry into the software business, 1950-1970



Source: 1970 Roster

Table 1. Top 20 regions in 1970 in terms of the number of software companies

3-DIGIT ZIPCODE	*3-DIGIT ZIP CODE AREA NAME*	*% OF ALL COMPANIES*
100	New York, NY	11.10%
606	Chicago, IL	4.30%
021	Cambridge, MA	3.80%
200	Washington, DC	3.00%
900	Los Angeles, CA	3.00%
191	Philadelphia, PA	2.30%
303	Atlanta, GA	1.90%
070	Fairfield, NJ	1.70%
554	Minneapolis, MN	1.70%
441	Cleveland, OH	1.50%
752	Dallas, TX	1.50%
943	Palo Alto, CA	1.50%
076	Paramus, NJ	1.30%
770	Houston, TX	1.30%
085	Princeton, NJ	1.10%
152	Pittsburgh, PA	1.10%
452	Cincinnati, OH	1.10%
481	Ann Arbor, MI	1.10%
802	Denver, CO	1.10%
850	Phoenix, AZ	1.10%

Source: 1970 Roster

Table 2. Top 20 regions in 1970 in terms of the number of software-industry employees

3-DIGIT ZIPCODE	***3-DIGIT ZIP CODE AREA NAME***	***% OF ALL EMPLOYEES***
900	Los Angeles, CA	19.90%
100	New York, NY	14.30%
904	Santa Monica, CA	7.00%
902	Redondo Beach, CA	5.80%
914	Sherman Oaks, CA	3.10%
068	Greenwich, CT	3.00%
076	Paramus, NJ	2.80%
117	Bethpage, NY	2.50%
554	Minneapolis, MN	2.40%
021	Cambridge, MA	2.20%
085	Princeton, NJ	1.80%
606	Chicago, IL	1.40%
200	Washington, DC	1.40%
752	Dallas, TX	1.40%
190	Philadelphia, PA	1.20%
037	Hanover, NH	1.10%
809	Colorado Springs, CO	1.10%
787	Austin, TX	1.10%
209	Silver Spring, MD	1.00%
944	San Mateo, CA	0.90%

Source: 1970 Roster

Table 3. Top 20 metropolitan regions in 1970 in terms of the agglomeration of software-industry employment

METRO AREA CODE	***METRO AREA NAME***	***TOTAL EMPLOYMENT***	***LOCATION QUOTIENT FOR SOFTWARE INDUSTRY***
5120	Minneapolis-St. Paul, MN	774,300	9.22
8840	Washington, DC/MD/VA	1,269,800	5.30
1920	Dallas-Fort Worth, TX	671,400	4.40
7400	San Jose, CA	422,300	3.69
5602	Bergen-Passaic, NJ	581,900	2.54
5600	New York-North-eastern NJ	3,671,700	2.37
8730	Ventura-Oxnard-Simi Valley, CA	139,800	2.35
3360	Houston-Brazoria, TX	809,700	2.33
4920	Memphis, TN/AR/MS	293,100	2.24
4120	Las Vegas, NV	113,200	2.17
0520	Atlanta, GA	602,000	2.04
2310	El Paso, TX	125,300	1.96
8480	Trenton, NJ	127,300	1.93
4480	Los Angeles-Long Beach, CA	2,879,100	1.91
3240	Harrisburg-Lebanon-Carlisle, PA	176,700	1.86
1760	Columbia, SC	136,700	1.80
9160	Wilmington, DE/NJ/MD	200,200	1.64
0760	Baton Rouge, LA	100,600	1.63
2080	Denver-Boulder-Longmont, CO	516,600	1.59
5603	Jersey City, NJ	259,100	1.58

Note # 1: Philadelphia ranked 25th with a location quotient of 1.35

Note # 2: Boston ranked 31st with a location quotient of 1.22

Note # 3: Chicago ranked 38th with a location quotient of 1.07

Source: 1970 Census

Table 4. Regression of location quotient for the software industry on location quotients for all high demanders (OLS estimates with robust standard errors between parentheses)

computer manufacturing	0.177
	(0.130)
accounting	0.065
	(0.140)
management	0.380**
	(0.130)
national security	0.157*
	(0.080)
banking	0.267
	(0.250)
insurance	0.108
	(0.150)
ordnance	0.028
	(0.030)
aircraft & parts	-0.027
	(0.030)
employment	0.334*
	(0.150)
employment squared	-0.012*
	(0.010)
constant	-0.339
	(0.230)
R^2	0.351
Degrees of freedom	117
Bayes Information Criterion	391.981

* = p < 0.05, ** = p < 0.01, *** = p < 0.001

Table 5a -- Individual regressions of location quotient for the software industry on location quotient for high demanders: computer manufacturers, accounting, management, and national security (OLS estimates with robust standard errors between parentheses)

	COMP MANUF	***ACCOUNTING***	***MANAGEMENT***	***NAT SEC***
computer manufacturing	0.186			
	(0.120)			
accounting		0.225		
		(0.130)		
management			0.564**	
			(0.180)	
national security				0.199*
				(0.100)
constant	0.488***	0.428**	0.255	0.429**
	(0.100)	(0.130)	(0.150)	(0.140)
employment	0.645***	0.662**	0.407*	0.712***
	(0.190)	(0.200)	(0.200)	(0.180)
employment squared	-0.023***	-0.023**	-0.014*	-0.025***
	(0.010)	(0.010)	(0.010)	(0.010)
R^2	0.238	0.109	0.185	0.150
Degrees of freedom	124	124	124	124
Bayes Information Criterion	378.513	398.522	387.189	392.482

* = p < 0.05, ** = p < 0.01, *** = p < 0.001

Table 5b -- Individual regressions of location quotient for the software industry on location quotient for high demanders: banking, insurance, ordnance, and aircraft and parts (OLS estimates with robust standard errors between parentheses)

	BANKING	***INSURANCE***	***ORDNANCE***	***AIRCRAFT***
banking	0.174			
	(0.220)			
insurance		0.117		
		(0.160)		
ordnance			0.073	
			(0.050)	
aircraft & parts				-0.011
				(0.020)
constant	0.484*	0.512**	0.556***	0.649***
	(0.200)	(0.170)	(0.090)	(0.090)
employment	0.679**	0.708**	0.726***	0.728**
	(0.240)	(0.210)	(0.210)	(0.220)
employment squared	-0.024**	-0.025***	-0.025***	-0.025***
	(0.010)	(0.010)	(0.010)	(0.010)
R^2	0.100	0.102	0.118	0.098
Degrees of freedom	124	124	124	124
Bayes Information Criterion	399.783	399.595	397.265	400.06

* = p < 0.05, ** = p < 0.01, *** = p < 0.001

Table 6. Regression of the location quotient for the software industry on the location quotient for one (combined) heavy-demander sector (OLS estimates with robust standard errors between parentheses)

	8 combined	*6 combined*
8 combined	0.624**	
	(0.230)	
6 combined		0.544*
		(0.220)
Constant	0.019	0.084
	(0.220)	(0.220)
employment	0.600***	0.632***
	(0.160)	(0.170)
employment squared	-0.021***	-0.022***
	(0.010)	(0.010)
R^2	0.176	0.167
Degrees of freedom	124	124
Bayes Information Criterion	388.517	389.869

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

"8 combined" includes all 8 heavy demanders, whereas "6 combined" excludes accounting and management

Table 7a. Regression of the location quotient for the software industry in 1970 on the location quotient for the combined high-demander sector in 1960 (OLS estimates with robust standard errors between parentheses)

	6 combined	***4 combined***
6 combined in 1960	0.722***	
	(0.123)	
4 combined in 1960		0.694***
		(0.117)
college 1970	0.358	0.316
	(0.310)	(0.283)
employment 1960	0.119	0.076
	(0.197)	(0.192)
employment 1960 squared	-0.002	0.007
	(0.033)	(0.033)
constant	-0.357	-0.281
	(0.349)	(0.329)
R^2	0.178	0.181
Degrees of freedom	54	61
Bayes Information Criterion	190.051	205.502

* = p < 0.05, ** = p < 0.01, *** = p < 0.001

"6 combined" excludes computer manufacturing and ordnance, which are not available in the 1960 Census, and "4 combined" further excludes accounting and management

Table 7b. Regression of the location quotient for the software industry in 1970 on the instrumented location quotient for the combined high-demander sector in 1970 (IV estimates with robust standard errors between parentheses)

	6 combined	***4 combined***
instrumented 6 combined in 1970	0.837***	
	(0.135)	
instrumented 4 combined in 1970		0.789***
		(0.129)
college 1970	0.391	0.326
	(0.294)	(0.272)
employment 1960	0.164	0.107
	(0.193)	(0.186)
employment 1960 squared	-0.010	0.003
	(0.032)	(0.031)
constant	-0.508	-0.383
	(0.341)	(0.321)
R^2	0.202	0.194
Degrees of freedom	54	61
Bayes Information Criterion	NA	NA

* = p < 0.05, ** = p < 0.01, *** = p < 0.001

"6 combined" excludes computer manufacturing and ordnance, which are not available in the 1960 Census, and "4 combined" further excludes accounting and management