Agglomeration Economies and the High-Tech Computer Cluster

Nancy E. Wallace and Donald W. Walls*

July 23, 2004,

Abstract

This paper considers the effects of agglomeration on the production decisions of firms in the high-tech computer cluster. We build upon an alternative definition of the high-tech computer cluster developed by Bardhan et al. (2003) and we exploit a new data source, the National Establishment Time-Series (NETS) Database, to analyze the spatial distribution of firms in this industry. An essential contribution of this research is the recognition that high-tech firms are heterogeneous collections of establishments. We explicitly model the kinship relationships between the headquarters and establishments of these firms and account for their endogenous production technology choices using controls for the spatial and functional configurations of each firm's establishment locations. The empirical results, from our preferred specification of a random parameters restricted maximum likelihood (REML) production function, are broadly consistent with several recent theoretical models of supply chain management under incomplete contracting (Combes and Duranton (2003), Almazan et al. (2003), and Rotemberg and Saloner (2000)). We find statistically significant and economically meaningful localization effects on high-tech firms' labor input technology arising from MSA-level proximity to workers in computer services Standard Industrial Classifications (SICs) and establishment-level geographic interactivity. We also find that localization effects have economically significant impacts on the elasticities of other purchased inputs. The channels for these effects are again access to large computer services labor markets and geographically dispersed networks of establishments. Our empirical results indicate that there are few benefits associated with firm locations in labor markets with large numbers of employees in the computer manufacturing sectors. This negative result may reflect the culmination of recent trends in out-sourcing manufactured inputs to distant offshore subsidiaries. Finally, we uncover considerable heterogeneity in the production technologies exploited by firms in the high-tech computer cluster, although in general, the production technology of this industry is characterized by constant returns to scale.

*Please address correspondence to (Wallace): Haas School of Business, University of California at Berkeley, Berkeley, CA 94720. Phone: (510) 642-4732. Fax: (510) 643-7357. E-Mail: wallace@haas.berkeley.edu.(Walls): Walls & Associates, Oakland, CA . Phone: (510) 763-0641. E-Mail: dwalls2@earthlink.net.

1 Introduction

A large body of economic research has focused on the role of agglomeration economies, or localized aggregate increasing returns, in the organizational structure of industries. Numerous causal mechanisms for their existence have also been suggested. In Marshall (1890), the benefits of agglomeration, or clustering, arise from the accumulation of human capital and productivity enhancements due to face-to-face communications. Krugman (1991) argues that the efficiency benefits of industry agglomeration exist because firms can better adjust their employment levels to respond to idiosyncratic productivity shocks. Helsley and Strange (1990) contend that labor market pooling reduces transaction costs associated with search and matching processes in the labor market. Localized technological spillovers have been identified as an alternative rationale for the existence of agglomeration efficiencies (See Fujita and Thisse (2002) and Fujita and Ogawa (1982)), however, strong empirical verification of these types of spillovers has been limited.[1] Generally, empirical tests of the effects of agglomeration economies must disentangle the effects of scale economies that are internal to the firm from the effects of external agglomeration economies arising from transportation cost advantages, intellectual or technology spillovers, or other innate natural advantages associated with given geographic locations.[2]

The purpose of this paper is to consider the role of agglomeration on the production decisions of firms in the high-tech computer cluster. We first build upon and extend an alternative definition of the high-tech computer cluster, or industry, developed by Bardhan et al. (2003). Given our definition of the industry, we exploit a new data source, the National Establishment Time-Series (NETS) Database, to analyze the spatial distribution of firms in the computer cluster. Using the NETS data, we define firms as sets of geographically dispersed establishments that report to a single headquarters. Each establishment, or plant, is identified by an eight-digit SIC code reflecting the attributes of its employees and outputs. The high-tech computer cluster is defined functionally by the activities of establishments, however, these establishments are, in turn, functionally interlinked to the firms that own them. We merge the NETS data with Compustat data, so that we can consider multi-factor specifications for the production technologies of publicly traded firms.

Our preferred empirical specification explicitly accounts for the coordination of production decisions between the headquarters and establishments of high-tech firms. It thus avoids potential biases in prior empirical analysis of this industry that do not account for the "jointness" of the production decisions of the many multi-establishment firms in the high-tech

[1] Jaffee et al. (1993) find a distance related decay function in the incidence of firm patent citations.

[2] In fact, Ellison and Glaeser (1997) argue that without firm specific information these effects cannot be separately identified.

computer cluster. Accounting for the internal structure of high-tech firms also allows us to differentiate the relative importance of localized aggregate increasing returns from scale economies that may arise from firms' internal production hierarchies. The idea behind this representation is that decisions made at headquarters are likely to affect all the production decisions of establishments that report to them. In addition, these decisions may be coordinated in order to derive spillover benefits from proximity to similar types of production units. We explicitly model the endogenous technology choices of high-tech firms using controls for the spatial and functional configurations of each firm's establishment locations. The state variables underlying the firm's endogenous technology decisions include measures for the composition of the local labor markets and the "within" firm establishment-level interactivity. We consider the effects of these characteristics on the production decisions of an unbalanced sample of the largest publicly-traded firms in the industry over a fourteen year period. Our estimation strategy provides tests for the economic and statistical significance of internal scale economies in production while controlling for the potential effects of localized aggregate increasing returns.

One advantage of the high-tech computer cluster is that this industry is less likely to benefit from innate natural resource advantages of location. Thus, the state variables underlying each firm's endogenous factor input and output choices, which we term the firm's technology choice, would be expected to include factors such as location specific human capital and the potential for knowledge spillovers. Of course, one challenge to this view of the firm's production decisions is that a firm's implementation of its technology is probably largely accomplished through the use of unobservable implicit and explicit contracts.

Several recent theoretical papers offer important insights into the location decisions of firms as mechanisms to solve the factor input supply problem under incomplete contracting (*i.e.* knowledge cannot be patented and exclusive labor contracts are typically not feasible). Almazan et al. (2003) identify conditions under which human capital is more efficiently created and better utilized in industrial clusters that contain similar knowledge-based firms. They conclude that firms would cluster when workers contribute to their own training and firm specific uncertainty is high. In contrast, growing industries in which firms invest substantial amounts in their workers' human capital would prefer separate locations.

Rotemberg and Saloner (2000) identify two key forces underlying regional agglomeration of production. The first is firm competition for the services of trained workers that assures that workers will earn a fair return on their industry specific human capital investments. The second is scale economies in production technology. Combes and Duranton (2003) assume that the propensity of workers to change jobs in the same local labor market is greater than their propensity to move between local labor markets. This assumption, in conjunction with incomplete contracting, implies that knowledge is partly embodied in workers, so poaching workers is a way for firms to raise their productivity. The costs of poaching include the risk that all proprietary productivity increasing knowledge is vulnerable and higher retention costs for key technology employees.

The results of these papers suggest that agglomeration leads to a trade-off between the benefits of labor market pooling and the costs of labor market poaching. Growing industries that need specialized firm specific training would choose to locate away from competitors due to their inability to appropriate knowledge from the training of competitors. In general, external increasing returns should favor agglomeration; however, crowding or market-impact effects should induce agents to prefer markets with fewer agents of their own type. These differing results suggest that the effects of labor market pooling on the efficiency of competitive firms is fundamentally an empirical question that requires a careful decomposition of the effects of scale economies internal and external to the firm.

The paper is organized as follows. We review some of the data problems that have plagued the empirical agglomeration literature in Section 2. Our functional definition of the computer cluster is developed in Section 3 and we use our definition to identify our sample of the largest publicly-traded firms in the industry. In Section 4, we consider the geographic agglomeration of the computer cluster and the spatial distribution of the largest firms. We also discuss the recent trends in the composition of employment both in the industry and in the sample. Section 5, introduces our strategy to estimate firm-level production functions for the sample of the largest publicly-traded firms and we describe the data used in this analysis. The estimation results are reported in Section 6 and we conclude in Section 7.

2 Empirical Agglomeration Studies

Most recent empirical studies of agglomeration economies have focused on industries defined by Standard Industrial Classifications (SICs) using plant-level data or geographic aggregates of plant-level employment. These studies rely on two primary data sources: the Census of Manufacturing and Dun and Bradstreet. The studies that rely on the Census of Manufacturing data are restricted to SIC codes that the Census Bureau defines as manufacturing. These studies must assume that modern U.S. industries can be correctly defined by aggregates of establishments that have like manufacturing SICs. An important disadvantage of this assumption is that there are many non-manufacturing SICs that are of increasing importance to the overall growth of the high-tech economy in the United States and yet are not included in the Census of Manufacturing. As will become obvious from our analysis, many key firms in the high-tech computer sector have well less than 50% of their employment in the manufacturing SIC of their headquarters; and that SIC is often the part of the firm that is growing the slowest (or even declining). Consequently, Ellison and Glaeser (1997) note that "the computer industry is a bit hard to find in the [manufacturing] SIC codes" (p. 902).

A second problem with the Census of Manufacturing source for these studies is that even within the manufacturing SIC codes, confidentiality restrictions and small sample sizes often limit the range of SIC codes that are available for the high-tech sector at county or MSA-level spatial units. For example, Henderson (2003) defines high-tech industries as: computers (SIC 357); electronic components (SIC 367); aircraft (SIC 372); and medical instruments (SIC 384); and excludes, because of small sample sizes: communications (SIC 381); search and navigation equipment (SIC 381); and measuring devices (SIC 382). Ellison and Glaeser (1997) identify the computer industry as electronic computers (SIC 3571); computer storage devices (SIC 3572); and semiconductors and related devices (SIC 3674); but then must use state-level aggregation to avoid confidentially restrictions. Moretti (2004) uses a definition of high-tech industries that he obtained from the American Electronic Association based on 45 four-digit SIC codes.[3] A key question is whether these definitions are useful to our understanding of the effects of agglomerative forces and spillovers on the productive efficiency of modern high-tech corporations that operate over many different SIC codes including many non-manufacturing codes.

Confidentiality restrictions also constrain researchers to trade-off geographic detail for SIC specificity. At smaller geographic designations, such as MSA's, nondisclosure requirements protect survey participants from identification at more precise higher-digit SIC codes. Greater functional precision for industries, say four digit SIC codes, is consistently available only for high-level geographic aggregates such as states. Most studies using aggregated geographic data have been limited to a stylized representation of firms' production technologies as a function of a single factor input, usually labor.[4] Three recent papers (See Dumais, Ellison, and Glaeser (2002), Henderson (2003), and Moretti (2004)) have used plant-level data obtained from the Longitudinal Research Data (LRD) base of the Census Bureau. These data allow for the examination of multi-factor production technologies including capital, labor, and materials at the two or three digit SIC level. Individual establishments are treated as autonomous decision-making entities that choose both their technology and the levels of all factor inputs. The LRD data provide indicator variables for establishments that belong to

[3]Moretti (2004) cites this definition to include "computers and office equipment, consumer electronics, communication equipment, electronic components, semi-conductors, industrial electronics, photonics, defense electronics, electromedical equipment, software and computer related services, and telecomunications services." (p. 21)

[4]For examples of these choices see Duranton and Puga (2001), Ellison and Glaeser (1997), and Glaeser et al. (1992) among many others, whose analyses are restricted to single-factor production functions.

multi-establishment firms, however, linkages in factors input decisions across establishments are not observable.

A third limitation with the data available from the Census of Manufacturing is that the data cannot be used to construct annual establishment-level or firm-level panels.[5] Since most econometric specifications for production functions rely either on fixed effects or first differenced estimation strategies, it is usually necessary that a establishment appears in the Annual Survey of Manufactures (ASM) in at least two different Census years. Since the ASM is only done in five year waves and the samples are different across these waves it is not possible to get a consistent year-by-year appraisal of the production decisions of establishments. Each five-year wave starts the second year after a Census and is drawn from the sample of establishments in the prior Census. Thus, an ASM wave covers a given year in terms of data, but it is picked from establishments that were drawn in the sample five years before. This sampling structure means that only larger establishments tend to be covered, because those are the only types that are included in every wave of the ASM. Although the ASM also includes a sampling of smaller establishments, that part of the sample changes with each wave so that the observed time lag between observations is often five years.

The second source of data is Dun and Bradstreet. These data identify the number of employment positions within each establishment defined by eight-digit SIC codes along with very detailed geographic information. Unfortunately, the output information (total sales in dollars) is only available at the headquarters-level[6], so the Dun and Bradstreet data are usually used to analyze the relative marginal product of labor within and across geographic areas (See Rosenthal and Strange (2001) and Henderson (2003)) within the framework of one factor input production technologies. An important, and as yet unexploited, advantage of the Dun and Bradstreet data is that they are collected from firms annually and they contain information on the firm-level ownership relationships among plants. These plant-level kinship relationships across SIC codes can be used to reveal the spatial distribution of establishment locations among competing firms in industries.

In summary, empirical analyses of agglomeration ideally should be carried out with as much geographic and functional employment activity detail as possible. In particular, it would be desirable to avoid aggregation problems that are likely to arise with the reliance on statistical summaries of industries that have been accumulated using administrative boundaries such as states or counties that may or may not correspond to the actual location patterns of co-located establishments (See Duranton and Overman (2004)). A further prob-

[5]This is because the full survey occurs only every fifth year and in the intervening years only selected "mini" panels are surveyed

[6]The sales for establishments are only reported in aggregate at the headquarters-level.

lem with past reliance on administrative spatial units is that they often vary in population and geographic area yet most prior studies treat these units as homogeneous. One would also want to control for the relationships between establishments in production rather than treating establishments as autonomous production units. Since many factor input decisions such as those related to physical or research and development capital are unlikely to be made at the plant level, it is important to control for the jointness of production costs across establishments in high-tech firms. Finally, the lack of high frequency panel data by firms and establishments has made it difficult to uncover important time series dynamics in location and production across firms.

3 Functional Definition of the Computer Cluster

There are two fundamental differences in the way this research defines the high technology computer industry: (1) we explicitly include services SICs in the definition of the computer sector and (2) we recognize that high technology firms are often made up of many establishments, not all of which may have as their primary activity what we think of as "high-tech." An essential hypothesis of this analysis is that, to understand key decisions made by these high tech firms, one must take into account the range of activities that the enterprise undertakes.

First of all, there is a surprising lack of consensus about the functional composition of the computer cluster in the United States. Following Bardhan et al. (2003) and the American Electronics Association,[7] our definition of the computer cluster includes *both* manufacturing (hardware) and service (software) components. We start identifying establishments in the computer cluster using both SIC codes. We identify firms in the computer cluster using both SIC codes and the new North American Industry Classification System (NAICS) coding formats. In Table 1, we summarize our definition of the high-tech computer cluster using both the NAICS and SIC coding systems.[8] We use both systems because in 1997, the NAICS codes became the official industry affiliation coding system, however, most empirical agglomeration studies have relied on the earlier SIC coding system. As shown, the manufacturing components of the industry include the high-tech electronics industry (NAICS 3341 and 3344); the semiconductor machinery manufacturing (NAICS 33295); and instruments for measuring and testing electronics (NAICS 334515).

Services production activities are also a significant part of the overall operation of many firms in the computer cluster. As shown in Table 1, our definition of the services component

[7] See $http://www.aeanet.org/Publications/IDMK-definition.asp$.

[8] Although not reported in Table 1 our functional breakdown is at the eight digit SIC code level.

of the high-tech computer cluster includes activities within the 5415 NAICS code; Computer Systems Design and Related Products. The functional production activities under this NAICS code include: computer system design and related services; software publishers; on-line information services; and data processing services. The NAICS/SIC codes associated with services activities have not been included in prior empirical agglomeration studies because these codes are not defined as manufacturing activities by the Census Bureau and thus do not appear in the Census of Manufacturing which is the primary data source used in these studies.

As shown in Table 1, several NAICS codes within our definition of the high-tech computer cluster do not exactly correspond to four digit SIC codes. This lack of correspondence between the two coding systems is a particular problem for NAICS 3329, Other Fabricated Metal Product Manufacturing. Semi-conductors can only be identified at the eight-digit SIC code level within this broad category of functional activity and similar problems arise within many of the broader NAICS designations. For these reasons, an accurate assessment of the computer cluster requires information on eight digit SIC codes and these data are not available through the Census for MSA-level geographic locations.

Our methodology for defining the high tech computer sector makes a second important distinction in that we identify the corporate relationships among all the establishments that define a "high-tech firm." In other words, our analysis also considers all the other establishments, including those whose primary activity is *not* covered in Table 1, that report to those establishments that we have identified as "high-tech." It is the totality of these linkages that defines the "high-tech firm;" and, we argue, this structure impacts the production decisions of those firms. Thus, when we define the high technology computer sector, it includes all establishments that either meet the criteria of Table 1 or report to an headquarters that does. This, of course, requires that we have access to a database that not only identifies the SIC activities of establishments, but also allows us to track corporate hierarchies over time.

3.1 Data for the Functional Analysis of the Computer Cluster

Our data set allows us to identify the functional composition of employment in the computer cluster using the SIC codes of establishments, the organizational relationship between establishments and firms, and the spatial distribution of competing firms. The National Establishment Time-Series (NETS) historical database provides time-series information on establishment mobility patterns, sales growth performance, job creation and destruction, changes in primary markets, and historical Dun and Bradstreet ratings. Walls & Associates constructed the database using fourteen annual snapshots of Dun and Bradstreet data and the full Duns Marketing Information (DMI) file. NETS follows over 27 million establishments between January 1990 and January 2003.[9] Because Dun and Bradstreet maintains policies and procedures regarding DUNS Numbers, the DUNS Number stays with an establishment even if the business unit relocates, merges or is acquired by another entity, has a management control or name change, or discontinues operation. This important feature of the data enables the NETS Database to track business dynamics at the establishment-level at eight digit SIC codes.

In order to systematically create a computer-related high technology sector, we first identified 401 8-digit industrial classifications (SIC8) in the Special Industry Machinery, Computer and Office Equipment, Telephone and Telegraph Apparatus, Electronic Components and Accessories, Instruments to Measure Electricity, and Computer and Data Processing Services sectors that met our standards for computer-related high-tech industries. We then searched the NETS Database, in every year 1989-2002, to find all establishments whose primary SIC8 was *ever* one of the 401 defining industries.[10] It is important to note here that the building block of our database is the "establishment," a business location with unique, separate and distinct operations.

Once we identified this core set of establishments, we then looked, in every year, for any establishments that reported to those establishments. The complete list of establishments (either originally identified or those reporting to an included establishment) represents the universe of the computer-related high technology sector. We then extracted annual information from the NETS Database for every year in which the uniquely identified establishments existed in the Database. While there were 296,026 establishments (260,835 firms) in our 2002 computer-related high technology sector, we tracked a total of 539,942 establishments during the period 1989-2002. Or, in other words, 45% of the establishments in this sector did not survive the entire period. In addition, in 2002, 7,978 of the firms were multiple-establishment firms.

Figure 1 summarizes the composition of jobs in the computer cluster in the United States from 1989 through 2002 using the NETS Database. Computer manufacturing SICs reported in Figure 1 359 (part), 357, 367, 3661, 3825, the Computer Services SIC is 737, and the

[9]Since the fourteen data poins are "snapshots" of the Dun and Bradstreet data as of January of every year 1990-2003, they actually refer to economic activity from the previous year. Thus, the panle of information covers 1989-2002 and we refer to these data throughout our discussion.

[10]Note that primary SICs (target markets) for establishments do change over time. So it is possible for an establishment to report its primary business as "Business oriented computer software" (SIC 73729902) and later-perhaps after a strategic refocus-report that its primary business line is "Compensation and benefits planning consultant" (SIC 87420201). This would indicate that the establishment shifted its focus from developing and delivering software to more consulting (where some software solutions might still be delivered). Since the establishment was in SIC 73729902 during the period, it would be included in the "high technology" universe throughout the period.

Other category is all other SICs. As shown, jobs in computer hardware manufacturing SICs have declined from about 35% of total high technology jobs to about 25% of total high-tech jobs in 2002. In contrast, the services components of the high-tech computer cluster have steadily grown over the same fourteen year period from less than 30% in 1989 to about 47% of total high-tech jobs in 2002. The level of the computer services component of the industry has steadily grown from an aggregate employment level slightly above 1,319,394 in 1989 to an employment level of about 2,646,324 employees in 2002. This change represents more than an 100% increase over the period. In contrast, manufacturing production activities have fallen steadily over the fourteen year sample period and total employment in these SIC designations was about 1,371,280 in 2002 representing about a 17% reduction. Figure 1 does not control for the number of firms, so it could just be that the services component of the computer cluster is growing because many new services companies and establishments are entering the market.

In Figure 2, we show the functional composition for the 177 largest publicly-traded firms in the computer cluster. Here again, the same trends in the relative importance of computer services and manufacturing components to overall U.S. employment are evident in the employment composition of the largest publicly-traded firms. Figure 2 reports the total number of jobs, recalling that Dun and Bradstreet surveys employment positions not employees, represented by the 177 largest publicly-traded firms and the percentage of those positions that fall within either computer manufacturing or computer services SICs. As shown in Figure 2, the computer manufacturing components of these firms have been steadily falling over the period, whereas, the services components have been steadily growing. In 2002, the manufacturing component of these firms had fallen to about 30% of overall employment and the services component had increased to about 30% of the total number of positions. The largest publicly-traded firms appear to be substituting away from computer manufacturing employment and by 2002 their overall employment had fallen. This trend has largely been missed in agglomeration studies, because prior studies have only focused on the computer manufacturing components of employment (See for example, Duranton and Puga (2001), Ellison and Glaeser (1997), and Glaeser et al. (1992) among many others), not the computer services components of this employment.

The importance of taking a broader view of the production activities of firms in the computer cluster is further illustrated in Tables 2 and 3. Using the NETS data, these tables summarize the functional employment diversity of large publicly traded high-tech firms in the computer cluster. With the exception of IBM and Motorola, the corporations in Table 2 self-identify themselves to Compustat as operating within the computer hardware SICs.[11]

[11]Phone conversations with analysts at Compustat revealed that firms are allowed to select their own

As shown in Table 2, there is considerable diversity both in the numbers of establishments operated by these firms and in the functional employment base of these establishments.[12] With the exception of Motorola and Lucent, all the firms have large numbers of jobs in the 357 SIC for computer manufacturing, however, this job component represents only 40% of Hewlett-Packard's jobs and and 30% of the overall jobs at IBM. Dell Computers, Cisco Systems and Seagate Computers have the highest proportion of SIC 357 jobs, yet all three firms have substantial numbers of jobs in other functions. The most important categories of jobs for these firms were SIC 737, Computer Services, SIC 504, Professional and Commercial Equipment, and SIC 87, Engineering, Accounting, Research, Management and Related Services.[13] The employment composition of Dell also has a very large component in management and public relations whereas the other firms have significantly fewer of these jobs.

The corporations in Table 3 self-identify themselves to Compustat as operating under the SIC 737 code, Computer and Data Processing Services. As shown, these firms tend to be heavily concentrated in SIC 737 designations, Microsoft and Novell report 80% and 84%, respectively, and BMC Software reports 100% of their employment in SIC 737. These services companies do have some level of employment in hardware functions as well as SIC, 504, Professional and Commercial Equipment. As is clear from both Tables 2 and 3, it is very unlikely that a sample of establishments that operate exclusively within high-tech manufacturing SICs would be representative of the population of high-tech *firms* in the computer cluster. Thus, inference based upon such samples should be viewed with some caution.

Another potentially important impediment to our current understanding of the production decisions of firms in the computer cluster arises from the importance of the spatial configuration of establishments within firms and the fact that these characteristics of firms have heretofore not been accounted for in the existing empirical literature. This problem is potentially most significant for studies that have focused on establishment-level data without controlling for the kinship relationships between plants and the likely "jointness" of production decisions across these establishments. Although it may be possible that establishment-level managers are allowed to control some parts of their factor input decisions, such as year-by-year employment levels, it is very unlikely that physical capital, research and development

overall SIC designation. Most of the firms in Tables 2 self-identify as operating within the 357 SIC, Computer and Peripheral Manufacturing Equipment. The 357 SIC is the most frequently used to represent the computer industry in empirical agglomeration studies.

[12] An establishment, in our data is equivalent to a plant in the Census LRD data only the NETS establishment data is identified for eight digit SIC codes.

[13] An separate analysis of the NBER Patent data for these firms reveals very large differences in number of scientific patents for Hewlett Packard and Cisco Systems, whereas all of Dell's patents are for business systems

capital, or "out-sourced" factor decisions are allowed at the establishment-level. These types of decisions are made at headquarters, as are all decisions concerning the geographic location of individual establishments within the firm. Treating individual establishments in multi-establishment firms as atomistic production decision makers is likely to lead to significant biases in econometric estimates of production technology if, in fact, the establishment-level production decision units are coordinated by headquarters. Finally, lack of controls for the kinship relationships between establishments and headquarters may lead to systematic underestimates of standard errors. Accounting for these relationships is necessary for unbiased inference as well as to unravel the relative importance of external spillovers in production from the internal structure of the firms' production technologies.

4 The Geography of the high-tech Computer Cluster

The NETS database allows us to identify the geographic locations for all establishments in our high technology computer sector. In Table 4, we report the relative geographic concentrations in 1989, 1995 and 2002 for the metropolitan areas with the largest agglomeration of computer cluster establishments. The fourteen year panel is comprised of any MSA that was among the largest 25 high tech employers in any year 1989-2002. This group of metropolitan markets accounted for between 55% and 58% of total sector jobs over the period. The number of computer sector jobs in these concentrated markets grew 41% from 1989 to the sectors peak in 2000; and, even after the sharp decline from that peak, grew 23% for the entire period. This average performance, however, masks considerable upheaval in the overall sector.[14]

Although San Jose and Washington, D.C. continue their number one and two ranking for overall employment in the computer high technology sector, San Joses jobs grew only 11% over the period while Washington grew 48%. There were also many significant changes in the relative rankings of MSAs over the fourteen year period. For example, Dutchess, NY, the home of IBM, lost half of its jobs over the period and its relative ranking fell from 21st to 58th place. Similarly, Colorado Springs (dominated by Digital Equipment in 1989) also lost one-half of its computer sector jobs by 2002 and fell from 23rd to 52nd among MSAs. On the other hand, Oakland, CA, with a dramatic 146% increase in sector employment over the period, improved its 1989 24th ranking to 10th by 2002. Likewise, Denver moved from 26th to 15th over the period, San Francisco nearly doubled its sector jobs to move from 19th

[14]Note also that only 28 MSAs account for all of the metropolitan markets that were ever in the "top25" during this period. Nonetheless, one would find it hard to conclude that agglomeration in these markets was anything but a dynamic process

to 13th; and Atlanta, with a 71% increase in jobs, moved from 12th to 7th among the most concentrated metropolitan markets.

To illustrate the significant dynamics of this sector, one out of four of the most concentrated MSAs over the period actually lost jobs in the computer sector despite the 23% growth in the group as a whole. Of those that expanded, the total change over the 14 years ranged from 7% (Baltimore, MD) to 146% (Oakland, CA) and was only mildly correlated with the ending size of the MSAs sector (0.14). Table 4 makes clear that the geographic locations of these establishments is quite dynamic and there are no clear patterns of either increased concentration or greater geographic dispersion in the computer cluster over the period 1989-2002.[15]

As is true in many industries in the United States, most of the employment and establishments in the high-tech computer cluster is accounted for by a relatively few large firms. We were able to obtain detailed Compustat data for two hundred and twelve of these firms, however, the effects of bankruptcy, mergers, and acquisitions reduced our analysis sample to 177 firms that had at least nine years of Compustat information from 1989 through 2002. In 2002, 166 of these firms were operating and they controlled 12,136 establishments in the U.S., and accounted for 1,754,556 domestic jobs. The NETS data provides the latitude and longitude for all establishments that reported to these firms and Figure 3 presents a map of these locations in 2002. As shown, there is quite wide geographic coverage of these establishments, although the heaviest concentrations appear in the largest metropolitan areas. The headquarter locations for the 166 largest publicly-traded firms are mapped in Figure 4 and this map demonstrates important concentrations of headquarters in the Bay Area and Los Angeles Basin in California, Chicago, Illinois, Austin, Dallas, and Houston, Texas, Boston, Massachusetts, and New York. The other important feature of Figures 3 and 4 is that they dramatically highlight problems that are likely to arise from the use of econometric specifications that erroneously assume establishments to be independent autonomous decision makers. These maps suggest that in 2002 there are one hundred and sixty six separate independent high-tech production decision units represented in these data even though there are 12,136 separate establishments. Figures 5 and 6 map the establishments for the manufacturing and software/service components of the high-tech computer cluster. Again, the concentration levels appear greatest in the most populous states, although there is more dispersion in the software/services establishment locations. Despite the apparent dispersion of these establishments.

[15] We also computed the Ellison and Glaeser indexes and their concentration ratios. These indexes exhibit remarkable little variation over the period. The indexes do not appear to accurately reflect either the industry concentration dynamics or changes in the geographic location of concentration in this industry over time. We hold a reconsideration of these indexes for future work.

One limitation of the mappings shown above, is that they do not provide a good representation of either the geographic organization of establishments within the largest publicly-traded firms or the proximity of each firm's establishments to the establishments of other firms in the cluster. We develop four measures to summarize the geographic network controlled by each headquarter. Our first two measures are intended to index the geographic dispersion of each firm's establishments from its headquarters. Using the longitude and latitude for each establishment, we compute the great circle mile distance between all establishments controlled by each headquarters. We define the *Inner Network Interactivity* of the firm as:

$$Inner\ Network\ Interactivity = \frac{\sum_i \sum_j w_{ij} \times e_i \times e_j}{\pi \times circ^2} \tag{1}$$

Where i and j are locations, w_{ij} is a weight assigned to the relationship between establishments i and j. The value of this weight is 1 for all relationships between the headquarters and the outlying establishments, .5 for all relationships between separate establishments, and zero for single establishment firms all within a 60 mile radius of the firm's headquarters. The e_i and e_j terms are the number of employees in establishments i and j respectively. The $circ$ term is the radius, 60 miles, of the "strong influence" interactivity of the firm's network of establishments.[16]

Similarly, we define our second measure, the *Outer Network Interactivity* for establishments that are located outside of the *Inner Network*, or greater than 60 miles from the headquarters of the firm,

$$Outer\ Network\ Interactivity = \frac{\sum_i \sum_j w_{ij} \times e_i \times e_j}{dist_{ij}^2} \tag{2}$$

where $dist_{ij}^2$ is the distance between locations i and j respectively. The w_{ij} weight is defined similarly to equation (1) for establishments farther that sixty miles from the firm's headquarters. The descriptive statistics for the establishment dispersion measures are reported in Table 5. The inner network interactivity measure ranges from 0, for single establishment firms, to 246,780 for firms with numerous large establishments within sixty miles of the headquarters. Because many of our firms have hundreds of establishments dispersed across the United States and large numbers of employees in each, our outer network interactivity measure ranges from 0, for single establishment firms, to 414,737,050 for firms with many large establishments that are more than sixty miles distant from the firm's headquarters.

We develop two further geographic measures to account for the composition of each firm's total labor market exposure in the MSA's in which it has establishments. These measures are

[16] This distance was chosen as a reasonable commuting distance for frequent face-to-face interactions.

intended to provide controls for the degree to which a firm's establishment location choices expose it to possible externalities from employees in the same SIC codes in the industry. The first of these measures is obtained using the NETS data to compute the firm's total MSA-level labor market exposure to workers in computer manufacturing SICs. The second measure is computed in the same way for the firm's total MSA-level labor market exposure to workers in computer services SIC codes. These measures sum the total labor market exposure for each firm over all the MSAs in which they have establishments.

The summary statistics for these measures are reported in Table 5. The largest publicly-traded firms in the industry locate their establishments so that on average they have access to 696,326 computer manufacturing SIC employees in the surrounding MSAs. The minimum is 423 for establishments that are located in a few smaller labor markets and the maximum is nearly the entire available labor market for these SIC codes. The average labor market access at the MSA-level is 473,402 for computer services employees. Here again, the largest firms have access to nearly the entire labor market through their establishment decisions. In summary, our geographic measures are intended to measure important state variables in each firm's production technology decisions. From the recent implicit contracting literature we would expect that decisions that are made at headquarters are very likely to affect a geographically dispersed set of production units and that the location of establishments is very likely to be coordinated with the co-location of similar types of employees who work in those MSA-level labor markets.

5 Firm Structure and the Estimation of Firm-Level Production Functions

Following Schoar (2003), Henderson (2003), and Moretti (2004), we assume that the production function for high-tech manufacturing firms can be approximated by a Cobb-Douglas production function in four inputs: labor; other purchased inputs; Research and Development Capital; and Physical Capital,

$$lnQ = ln\alpha_0 + \sum_i \alpha_i lnX_i + \epsilon. \quad (3)$$

The lnQ is the log of outputs and lnX_i is the log of fixed and variable factor inputs. The ϵ is an error term comprised of three components: factors that are known to the producer and affect production decisions; factors that are known to the producer and affect the *ex post* production decisions but are not predictable by the producer *ex ante*; and factors arising from measurement error or data collection problems that do not affect the producers' decisions. As is well known, the Cobb-Douglas production function imposes several important restrictions including unitary elasticities of substitution, constant production elasticities, and constant factor demand elasticities. For this reason, we also consider a translog functional form as a generalization of the Cobb-Douglas.[17]

The translog is an attractive functional form because as noted by Fuss et al. (1978) it provides the minimum number of parameters needed to represent economic behavior without imposing arbitrary restrictions on that behavior. We estimate a four input translog of the form:

$$lnQ = ln\alpha_0 + \sum_i \alpha_{i0} lnX_i + (.5)\sum_i \sum_j \alpha_{ij}(lnX_i)(lnX_j) + \epsilon^*, \tag{4}$$

where i indexes the factor inputs $i = 1, ..., N$ and j indexes the factor inputs, $j = 1, ..., N$. Since the Cobb-Douglas is nested in the translog, it allows us to test for the translog functional representation of the production elasticities.

Following Marschak and Andrews (1944) and Griliches and Mairesse (1995), it is likely that variable factor inputs are co-determined by the output objective function of the firm's decision makers. For this reason we assume that firm's make endogenous "technology" choices concerning the contemporaneous mix of variable factor inputs and outputs. These choices of "production technology" are determined by market determined state variables that are exogenous to the firm's output decision. Not recognizing that the endogeneity of these technology decisions, would of course lead to biased estimates of the true structural parameters of the production function. Our strategy to estimate this endogenous technology choice is described in the next two subsections.

5.1 Data

To obtain our analysis sample, we identified headquarters in our high technology universe that had income statement data available from COMPUSTAT. In every year, we isolated those establishments that reported to the target headquarters. They became the basis of our empirical analysis. We ultimately focused on an unbalanced panel of 212 firms of which 177 firms had operated for at least nine consecutive years from 1989 through 2002. They included 13,303 establishments and 1.75 million jobs or 32% of the overall sector in 2002 (4% of establishments). Our method to develop the panel does, of course, introduce the possible effects of survivorship bias even among the largest firms due to bankruptcies, mergers, and

[17] The translog is obtained from the Cobb-Douglas by specifying the Cobb-Douglas production elasticities to be log-linear functions of inputs. That is by adding $\alpha_i = \alpha_{i0} + (.5)\sum_j \alpha_{ij} lnX_j$ to Equation 3 above.

acquisitions.[18]

For the sample of 177 large publicly traded companies, we obtained COMPUSTAT income statement data to construct each firm's factor input measures for labor, other purchased inputs, the stock of net research and development capital, and the stock of net capital. The output measure is total sales (total value of shipments) plus changes in the value of inventories for finished goods and work-in-progress. Ideally, our production functions would be estimated using an output measure for the actual quantities of outputs. Clearly, if product markets are not perfectly competitive, the estimated residuals from the production function regressions may reflect variations in efficiency in addition to differences in mark-ups. Unfortunately, data availability restricts our choice of output variables, so we follow other authors and use total sales (See, Schoar (2003) Henderson (2003); and Moretti (2004)).

We measure labor inputs as the total number of employees using information from COMPUSTAT. Our measure of other purchased inputs was computed using the COMPUSTAT income statement measure of the "Cost of Goods Sold" (CGS). CGS is a very significant expenditure category that includes expenditures for parts, intermediate goods, fuel and energy purchased, and inputs from contracted work. Unfortunately, a significant weakness of these data is that high-tech firms also include their labor and related expenses in "Cost of Goods Sold". To compute net CGS, we estimated the firm's total wage bill using *County Business Pattern* data to measure the average MSA-level wages for each establishment's type of SIC employment. The firm's average wage rate was computed as a weighted average of its establishment-level wages. The Country Business Patterns data allows us to calculate a weighted average wage rate in 1992 and 1997, we then estimate the five year change in the firm's wage rate to impute an annual time series of average wage rates for each establishment. We then multiplied the weighted average wage rate by the numbers of employees and subtracted this estimated total wage bill from CGS. Values for the capital stock were measured using the COMPUSTAT income statement measure for "Net Purchased Plant and Equipment" for the net book value of capital. We then applied the perpetual inventory method used in Schoar (2003). We compute our measure for Research and Development Capital using COMPUSTAT income statement reports on research expenditures and then applied the perpetual inventory method outlined in Hall and Mairesse (1995).[19]

Summary statistics for our input and output measures are reported in Table 6. As shown,

[18] A good rule of the thumb for most sectors-high tech or not-is that about one-half of all establishments that existed over the 1989-2002 period are not in business at the end of the period. We plan to address these potential problems in future work.

[19] We assumed a depreciation rate of 15 percent, a pre-sample growth rate of 5 percent in real research and development expenditures, and we start the recursive perpetual inventory formula at least three years before the sample period.

the firms in the sample are quite large. The mean number of employees is 12,018 and the maximum number is 383,220 employees. The average number of establishments operated by these firms is about 73 and the largest publicly-traded firm operates 637 establishments. The average depreciated book value of the capital stock is about $556.25 million in 1989 dollars and the largest publicly-traded firm controls a capital stock of over $26 billion. There was considerable variation in the research and development stock of these firms. The average depreciated book value of R&D was $850.45 million, however, a number of firms had no R&D capital stocks. There is considerable dispersion in our computed measure of other purchased inputs. The average value of this factor input is $1.5 billion and the standard deviation is $4.2 billion.

The important benefit of our merging the geographically rich NETS data and the COMPUSTAT data is that it allows for richer representations of factor input decisions including the jointness in production for factors such as capital and R&D expenditures. Our measures for the establishment-level kinship relationships and the geographic location of the establishment networks provide potentially very powerful controls for the effects of internal and external labor market spillovers in the production technology of these firms. Similar controls have not been exploited in this way in prior empirical agglomeration research.

5.2 Production Function Results

We hypothesize that the major misspecification which is transmitted to the factor decisions, such as differences in land or entrepreneurial and labor quality, are essentially fixed over time and can be eliminated with a ”within” transformation controlling for ”fixed” individual firm-level effects and time effects. In addition, we assume that our measures of the geographic dispersion of the firms establishments and the labor market structure of their chosen establishment locations are fixed *ex ante* to production decisions. We justify this assumption, because the complicated locational structures of these firms reflect contractual commitments in the form of leases or property purchases. Since real estate is notoriously illiquid, commercial lease contracts have maturities in the five to seven year range, and industrial lease contracts are structured in the seven to ten year maturity range, we feel that the overall structure of the individual firms location decision is justified. The form of our Cobb-Douglas specification is thus:

$$lnQ = ln\alpha_{i0} + \sum_i \alpha_i lnX_i + \sum_k \gamma_k lnZ_{k,t-1} + \epsilon. \tag{5}$$

The lnZ_k is a vector of lagged exogenous firm establishment network structure and labor market access variables: the natural log of *Inner Network Interactivity*, the natural log of

Outer Network Interactivity, the natural log of each firm's access across all of its establishments to total MSA employment in computer manufacturing SICs and the natural log of each firm's access across all of its establishments to total MSA employment in computer hardware SICs. We report the results the parameter estimates of our "fixed effects" and Huber-White standard errors in Table 7 for equation 5 and for the translog version of the same specification. All the right -hand side variables have been deflated to 1989 dollars and then de-meaned, using the firm specific mean. We also include included time dummies which we do not report. As shown, in Table 7, the coefficients on the factor inputs all have the expected sign, however, the coefficient on R&D Stock is extremely small and is not statistically significantly different from zero. Our *Inner Network Interactivity* measure has a statistically significant and positive effect on output suggesting that firms with many establishments within a sixty mile radius of the headquarters of the firm have production advantages to firms with establishments that are more dispersed. Firms located such that their establishments have the greatest access to Computer Manufacturing SIC employees appear to be at a production disadvantage relative to firms that locate their establishments such as to afford access to MSA's with large number of Computer Services SIC employees. The effects of both of these MSA-level labor market access variables are statistically significant at better than the 5% level.

The importance of the labor market access effects is also economically more significant than the importance of the within-firm establishment interactivity effects. Our estimation results suggest that there are potentially important "spillovers" through the channel of the firm's establishment network's proximity to employees that work in similar services SIC technologies. This result is contrary to a recent finding in Henderson (2003) who found that establishments belonging to multi-unit establishments depended more on internal-firm networks and were more insulated from local external environments than single establishment firms. These results were based on establishment-level production functions using on Census of Manufacturing SIC codes. We find that there are significant diseconomies associated with establishment networks that provide high levels of exposure to computer manufacturing SIC codes.

The translog results suggest that there are statistically significant non-linearities in the production functions. Our test of the Cobb Douglas restriction was rejected at better than the .001 level of statistical significance.[20] The coefficient estimate on the other purchased inputs is large and statistically significant, which may be indicative of current trends in "outsourcing" inputs among high-tech firms in the computer cluster. Further evidence of these

[20]The statistic was ($F_{10,2152} = 5.29$), for the test of the null hypothesis that all the parameter estimates on the squared and interaction parameters are jointly zero

trends is the statistically significant and negative interaction term between the employment factor and other purchased inputs. As is frequently the case, using fixed effects models with time interactions, the corrections often lead to low parameter estimates for the R&D coefficients. Overall, the results indicate that the underlying technology for high tech computer cluster firms is slightly decreasing returns to scale. We choose not to overly interpret the results reported in Table 7 due to our concerns that the specification does not adequately handle the likely endogeneity of firm-level technology.

6 Accounting for Heterogeneity in Firm-Level Technology

As previously discussed, equation (5) could still be subject to important remaining identification problems primarily due to the potential endogeneity of the variable factors. As noted by Griliches and Mairesse (1995), the within transformations implemented above are unlikely to correct for all simultaneity problems because they do not completely correct for firm-level heterogeneity in production. To address the potential differential effects of the firms' *ex ante* establishment network decisions, we follow an idea proposed by Mundlak (1988) and specify the coefficients of the variable factor as functions of variables reflecting inter-firm heterogeneity in technology or location pre-conditions.

Following Mundlak and Hellinghausen (1982), we assume that variable input technology is defined as a collection of techniques and each technique is represented by a production function. Technology is a population of production functions so that the Cobb-Douglas production technology can be specified for the jth firm as:

$$\mathbf{y}_j = \alpha_0 + \mathbf{X}_{Fj}\alpha + \mathbf{X}_{Vj}\beta_j + \varepsilon_j, \tag{6}$$

where

$$\varepsilon_j \sim N(\vec{\mathbf{0}}, \sigma_j^2 \mathbf{I}_{n_j}),$$

n_j is the number of observations (years of data) on the jth firm, σ_j^2 is the error variance associated with the $n_j \times 1$ vector ε_j, β_j is a $K \times 1$ vector of coefficients hypothesized to vary randomly across firms (but to be stationary within a firm), $\mathbf{X}_{Vj}$ is the $n_j \times K$ matrix of observations on the K variable factor inputs for the jth firm, $\mathbf{y}_j$ is the $n_j \times 1$ vector of observations on the output level of the jth firm, and $\mathbf{I}_{n_j}$ is an $n_j \times n_j$ identity matrix. The assumption of a common technology pool implies that all the β_j are drawn from the same population. However, the choice from the collection is not completely a random drawing. Instead, the coefficients on the variable factors are themselves functions of exogenous state variables representing the firm's network interactions and global exposure to local labor markets arising from the spatial location of all its establishments. Let $\mathbf{Z}_{j,t-1}$ be the $K \times L$ block diagonal matrix of the firm's aggregate $t-1$, *ex ante*, decisions concerning the aggregate spatial configuration of all the firm's establishments,

$$\mathbf{Z}_{j,t-1} = \begin{pmatrix} \tilde{\mathbf{z}}'_{j1} & \vec{\mathbf{0}} & \vec{\mathbf{0}} \\ \vec{\mathbf{0}} & \ddots & \vec{\mathbf{0}} \\ \vec{\mathbf{0}} & \vec{\mathbf{0}} & \tilde{\mathbf{z}}'_{jK} \end{pmatrix},$$

where $\tilde{\mathbf{z}}_{jk}$ is an ℓ_k-element vector of variables hypothesized to determine the network technology effects on the firm's variable input choices. Then the following set of equations determines the random coefficients β_j:

$$\beta_j = \mathbf{Z}_{j,t-1}\delta + \omega_j, \tag{7}$$

and

$$\omega_j \sim N(\vec{\mathbf{0}}, \mathbf{\Omega}),$$

for firms $j = 1, \ldots, J$ independently. δ is a fixed $L \times 1$ vector of aggregate *ex ante* localization effects and $\mathbf{\Omega}$ is a fixed $K \times K$ disturbance covariance matrix for these effects.

In what follows, we assume that the ε_j and ω_j are uncorrelated. This does not seem unreasonable given their different origins: The ω_j represent *ex ante* errors in lease contracting for the aggregate network of establishments, while the ε_j are *ex post* shocks to the firms output production decisions arising from non-contracted for or extra-contractual events.

If (7) is valid, then the firm-level β_j's are *random* coefficients drawn from a normal distribution of population parameters centered at $\mathbf{Z}_{j,t-1}\delta$. Ordinary least squares estimation does not account for variance component structure, $Var(y_j)$, which is a function of both ε_j and ω_j. Thus, it is an inefficient estimator and the standard errors of the estimates would be biased upward (See, Hsiao (1986), Laird and Ware (1982)). In addition, Stein-like estimators have been shown to be a superior method for incorporating prior structural information, such as equation (7) (See, Dempster et al. (1981)) .[21]

Because we're interested in (i) obtaining unbiased and efficient estimates for the β_j and

[21] Mundlak (1978) argues that the individual effects should always be treated as random effects and that the fixed-effects model should always be analyzed conditionally on the effects present in the observed sample.

δ parameters and (ii) drawing population inferences, equation (6) is more correctly viewed as a random-coefficients model in which the regression coefficients are assumed to be the dependent variables of another regression such as equation (7). Combining these two equations yields a random-coefficients model (general hierarchical model) for each firm, j, in our panel of high-tech firms in the computer cluster:

$$\mathbf{y}_j = \mathbf{X}_j \mathbf{Z}_{j,t-1}\delta + (\varepsilon_j + \mathbf{X}_j \omega_j)\,. \tag{8}$$

Equation (8) allows for firm specific heterogeneity in the firm's choice of production function as determined by its contractual choice of localization effects. The relationship between the locational and network externalities of these firms and their factor input choices can be tested using the estimates of δ. The model also provides estimates of the firm-level elasticities themselves, β_j, and their standard errors. Hence, it allows for direct tests of overall establishment location effects on the technological production decisions of these firms.

As previously noted, Equation (8) cannot be estimated by ordinary least squares (equivalently, *un*restricted maximum likelihood) because the combined error term $(\varepsilon_j + \mathbf{X}_j \omega_j)$ is correlated with the independent variables, unless $E\{\mathbf{X}_j \omega_j\} = \vec{\mathbf{0}}$. The consequence of this correlation is to bias downward the estimates of the variance-covariance matrix of ε_j and ω_j obtained from (unrestricted) maximum likelihood estimation. We can avoid this bias by using a restricted maximum likelihood (REML) estimator that accounts for the loss in degrees of freedom from estimating δ

We further assume that all firms in our sample respond to the same set of locational variables and that, within a firm, each contract coefficient (e.g., β_{jk}) is determined by a common vector of locational effects representing the firms network interactions and local labor market exposure (our estimation strategy would, however, allow us to relax this assumption). That is, $\tilde{\mathbf{z}}_{jk} = \tilde{\mathbf{z}}_{jn}$ for all $k, n \in \{1, \ldots, K\}$.

The results for our REML estimates of the Cobb-Douglas production function are reported in Table 8. The elasticity of the capital stock input is considerably reduced in the random parameters model compared to the results in either the Cobb-Douglas (equation (6)) or the Translog (equation (4)). The elasticity of the R & D stock has increased and is statistically significantly different from zero at better than the .001 level of statistical significance. A one percentage point increase in a firm's capital stock increases output by about twelve percentage points, whereas, a one percent increase in R & D stock increases output by only about one percent. The reduction in the overall estimated elasticity of the fixed factor inputs in the production technology of firms in the high-tech computer cluster suggests that coefficient estimates for fixed parameters production function specifications, even with flexible functional forms, ascribe overly high scale effects to the fixed factor inputs. Opening a channel for scale effects, that are external to the firm though labor market localization and firm network interaction effects and accounting for the affect of these state variables on the variable factor input techniques that are uniquely chosen by each firm, appears to importantly diminish the scale effects internal to the firm through the effects of capital structure on production.[22] A likelihood-ratio test reveals that the exogenous factors are jointly significant at better than the 1% level.

As shown in Table 8, the *ex ante* localization effects of establishment interactions and exposure to MSA-level labor markets have statistically significant effects on the variable factor elasticities of the largest publicly-traded firms in the high-tech computer cluster. Our finding that these exogenous factors are statistically and economically important in the determination of the variable factor elasticities supports our contention that different high-tech firms in the computer cluster employ different technologies and that these differences across firms are important. Firms with higher levels of employee interactions (*e.g.* firms with less dispersed networks of establishments) increase the elasticities of both the labor and other purchased inputs. The localization effect derived from having establishments that are close to large agglomerations of computer manufacturing SICs does not have statistically significant effect on the variable labor input elasticity. However, the localization effects of exposure to competitors' employees in computer services SICs is statistically significant at better that .1%. A one percent change in this localization effect increases the firm's variable labor elasticity by about thirteen percentage points. Overall, these effects suggest that the global effects of localization externalities and within firm network interactions lead firm decision makers to heavily weight their internal labor factor relative to firms that do not have these characteristics in their geographic structure.

The results for the elasticity of other purchased inputs are somewhat different. Here again, the level of exposure to the computer services MSA-level labor markets and the network interaction effects are all statistically significant at standard levels (although the outer network interactivity is only significant at the ten percent level). Interestingly, the more locational dispersion in the geographic location of the establishment structures the greater the weight the firm decision makers place on the use of other purchased inputs. The other purchased input elasticity is not statistically or economically sensitive to increased access to MSA-level labor markets in computer manufacturing. Whereas access to large MSA-level labor markets in computer services SICs decreases the firms technological dependence on other purchased inputs. The localization effects of exposure to competitors' employees

[22] Although not reported here, a specification that included random coefficient estimates for the fixed factors showed that the R&D and capital stock coefficients are not functions of the underlying state variables

in computer services SICs is statistically significant at better that .1% and a one percent change in this localization effect decreases the firm's variable other purchased input elasticity by about twelve percentage points. This localization effect appears to have an economically important and positive effect on the productivity of firms in the high-tech computer cluster.

Given our prior discussion, recent trends in this industry indicate that firms have significantly reduced their relative use of manufacturing labor and increased their relative use of computer services labor over the last fourteen years. The evolution of these state variables suggests that the relative scale effects of other purchased inputs is reduced with manufacturing employment because as shown by Bardhan et al. (2003) most of the recent production efficiencies in the high-tech computer cluster have been achieved through purchasing other factor inputs in the form of mass produced manufacturing components from distant, off-shore producers, rather than from the localized manufacturing factor input markets. These recent trends appear to imply that the wage increase or congestion effects of proximity to large numbers of competitors' employees in computer manufacturing outweigh the potential advantages of these localization effects through increased labor market availability.

Overall, the localization effects of proximity to manufacturing and computer services employment and the network interaction effects of firms' choices concerning the staffing and spatial location of establishments lead to increases in the mean elasticities for both the labor and other purchased inputs. The REML coefficient estimates can be used to generate distributions of parameters for both the variable employment and other purchased input factors. As shown, in last two lines of Table 8, the mean elasticity for the variable employment factor input is .273 with a standard deviation of .319 and the mean elasticity for the variable other purchase inputs factor is .65 with a standard deviation of .298. In Figures 7 and 8 we plot the histograms for the firm-level estimates of the employment and other purchased input elasticities. As is clear from the plots, there is considerable heterogeneity in the firm-level technology (the relative elasticity of employment and other purchased inputs) chosen by high-tech computer firms. The coefficient distributions are slightly more skewed and have fatter tails than would be expected in normal distributions, however, they are generally symmetric. The standard errors of these sampling distributions are .023 and .022, respectively, indicating that we would strongly reject the null hypothesis that these elasticities were zero at better than the .001 level of statistical significance. We plot the co-variation of the REML estimates for the firm-level employment elasticity and other purchased inputs elasticity in Figure 9. As shown in Figure 9 there is a lot of heterogeneity among firms in their relative elasticities of employment and other factor inputs. Firms with higher elasticities of the employment factor tend to have lower elasticities of other purchased inputs, and *vice versa*, clearly indicating the significant substitutability of these factor inputs.

The firm-level returns to scale can be computed from the production function as,

$$S = \alpha_{R\%D\ Capital} + \alpha_{Physical\ Capital} + \beta_{Employment,i} + \beta_{Other\ Purchased\ Inputs,i} \tag{9}$$

where the $\alpha_{R\%D\ Capital}$ and the $\alpha_{Physical\ Capital}$ are the REML estimated elasticities on the fixed factor inputs and the $\beta_{Employment,i}$ and $\beta_{Other\ Purchasd\ Inputs,i}$ are the REML random parameter estimates for the i^{th} firm's variable factor elasticities. As shown in Figure 10, there is considerable heterogeneity in the level of scale economies in the computer cluster. In contrast to the fixed parameters specification, the REML results suggest that the overall production technology of the high-tech computer cluster is characterized by constant returns to scale with a mean sum of the factor elasticities of 1.21 and a standard deviation of .14. As shown, the sampling distribution of returns to scale for the computer cluster is generally symmetric, however, the upper tail of the distribution is somewhat right skewed. It is particularly instructive to note, that firms such as Microsoft show up in the upper tail of the estimate of overall scale economies with a value of 1.45. Other firms that appear in the upper tail of the distribution include BroadVision Inc., Komag Inc., Incyte Corporation, and General Magic Inc. Firm that locate in the decreasing returns tail of the distribution include: Tarantella Inc., Xircom Inc., Centura Software and Computer Associates. In 2002, Hewlett Packward had absorbed Compaq Computers and, as the acquiring firm, HP has an estimated scale elasticity of 1.028.

The firm-level distribution of returns to scale, is in part determined through the channels of the localization externalities afforded by the proximity of computer services SIC labor employment and the network interaction effects of the spatial location of each firm's configuration of establishments across the United States. Our statistical results appear to suggest that these channels directly affect the production technologies and factor choices of the largest publicly traded firms in the high-tech computer cluster. An important advantage of our REML estimator is that it allows for a decomposition of the relationship between scale economies that are internal to the firm and those that arise from external factors.

7 Conclusions

In this paper we consider the possible effect of agglomeration on production decisions in the high-tech computer cluster. We build upon an alternative definition of the high-tech computer cluster developed by Bardhan et al. (2003) and we exploit a new data source, the National Establishment Time-Series (NETS) Data, to analyze the spatial distribution of firms in the computer cluster. The NETS Database allows us to identify firms as sets of establishments, that produce within specific SIC codes and report to single headquarters. We merge the spatial NETS data with Compustat data, so that we can consider richer specifications of the production technologies of firms in the computer cluster including controls for the spatial and functional configurations of the firms' establishment locations. We consider the effects of these characteristics on the production outcomes over a fourteen year panel of the largest publicly-traded one hundred and seventy seven firms in the industry. Our estimation strategy provides tests for the economic and statistical significance of scale economies in production controlling for the effects of local labor localization externalities and establishment network interactivity. Our econometric results appear to verify the importance of considering richer production function specifications that control for the jointness of capital and R & D factor input decisions in production.

The empirical results from our preferred specification of a random parameters (REML) production function are broadly consistent with the several recent theoretical models developed by Combes and Duranton (2003), Almazan et al. (2003), and Rotemberg and Saloner (2000). Consistent with all three models, the localization effects of proximity to MSA-level labor markets comprised of large numbers of computer services SIC employees positively affects the elasticity of the labor and other purchased inputs variable factors. These benefits can variously be explained by the insurance effects of a large available labor market pool or the possible efficiency gains arising from poaching specialized talent from competitors. The countervailing effects of the diminished network interactions from broad exposure to many labor markets appears to offset these efficiency gains because our results indicate that smaller more concentrated networks of establishments are more efficient. None of the recent theoretical literature considers, localization effects of labor market externalities on other variable factor inputs. Our results suggest that labor market localization externalities increase the efficiency of other purchased inputs only for proximity to computer services labor markets. Recent trends in outsourcing manufactured inputs to distant subsidiaries have reversed the benefits of locating near labor markets with large numbers of competitors' employees in computer manufacturing SICs. Efficiency gains for this factor input are available from widely dispersed establishment networks and proximity to competitors employees in computer services SICs. Finally, we uncover considerable heterogeneity in the production technologies exploited by firms in the high-tech computer cluster, although in general, the production technology of this industry is characterized by constant returns to scale.

Table 1: Functional Composition of the Computer High-Tech Industrial Sector

NAICS Industry Name	NAICS Code	SIC Code Equivalents	SIC Industry Equivalents
Other Fabricated Metal Product Manuf.	**3329**		
Semiconductor Machinery Manuf.	33295	35599927	Semiconductor Manuf. Machinery
		35599939	Electronic Comp. Making Machinery
Computer and Peripheral Manufacturing Equipment	**3341**		
Electronic computer Manuf.	334111	3571	Electronic Computers
Computer Storage Device Manuf.	334112	3572	Computer Storage Devices
Computer Terminal Manuf.	334113	3575	Computer Terminals
Other Computer Peripheral Equip. Manuf.	334119	3577	Computer Peripheral Equip.
		3578	Calculating and Accounting [1.]
Semiconductor and Related Manufacturing	**3344**		
Electron Tube Manuf.	334411	3671	Electron Tubes
Bare Printed Circuit Board Manuf.	334412	3672	Printed Circuit Boards
Semiconductor and Related Device Manuf.	334413	3674	Semiconductors and Related Devices
Electronic Capacitor Manuf.	334414	3675	Electronic Capacitors
Electronic Resistor Manuf.	334415	3676	Electronic Resistors
Electronic Coil, Transformer,	334416	3677	Electronic Coils, Transformers,
etc. Manuf.		3661	Telephone Apparatus [1.]
Electronic Connector Manuf.	334417	3678	Electronic Connectors
Printed Circuit Assembly Manuf.	334418	3679	Computer Peripheral Equip. [1.]
		3661	Telephone Apparatus [1.]
Other Electronic Component Manuf.	334419	3679	Electronic Components [1.]
Navigational, Measuring, Electromedical, and Control Instr.	**3345**		
Instruments for Measuring and Testing of Elec.	334515	3825	Instruments for Measuring and Testing of Elec. [1.]
Computer Systems Design and Related Products	**5415**		
Software Publishers	511210	7372	Software Publishers
On-Line Information Services	514191	7375	Informational Retrieval Services
Data Processing Services	514210	7374	Data Processing and Information
Computer Program Services	541511	7371	Computer Programming Services
Computer Systems Design Services	541512	7373	Computer Systems Design
		7379	Other Computers [1.]
Computer Facilities Services	541513	7376	Computer Facilities Services
Other Computer Related Services	541519	7379	Other Computer Services [1.]

Source: US Census Bureau, "Bridge Between NAICS and SIC," EC97X-CS3 (June, 2000)

[1.] Only a portion of the SIC code is included in the corresponding NAICS Code

Table 2: Functional Job Composition in 2002: High-Tech Hardware Firms

	Hewlett Packard	IBM	Cisco Systems	Dell Computers	Apple Computer	Intel	Seagate Technology	Motorola	Lucent Technologies
(SIC) Codes	%Jobs	% Jobs	% Jobs	% Jobs	% Jobs	% Jobs	% Jobs	% Jobs	
349 Misc. Fabricated Metal Products	0%	0%	0%	0%	0%	12%	0%	0%	0%
357 Computer and Office Equipment	40%	30%	62%	61%	47%	32%	86%	3%	11%
366 Communications Equipment	0%	1%	4%	0%	0%	0%	0%	31%	39%
367 Electronic Components, Access.	2%	5%	0%	0%	0%	37%	0%	26%	11%
369 Misc. Electrical Equip., Supplies	2%	0%	0%	0%	3%	0%	0%	4%	0%
382 Measuring and Controlling Devices	3%	0%	0%	0%	0%	3%	0%	0%	0%
384 Medical Instruments and Supplies	3%	0%	0%	0%	0%	0%	0%	0%	0%
422 Public Warehousing and Storage	1%	0%	0%	0%	0%	0%	0%	0%	1%
481 Telephone Communication	0%	0%	0%	0%	0%	0%	0%	5%	5%
504 Professional Commercial Equipment	18%	27%	7%	0%	13%	0%	0%	1%	0%
506 Electrical Goods	4%	0%	0%	0%	0%	0%	0%	5%	4%
573 Radio, Television, Computer Stores	2%	2%	2%	3%	2%	0%	0%	0%	0%
596 Nonstore Retailers	0%	2%	0%	0%	0%	0%	0%	0%	0%
599 Retail Stores, NEC	0%	4%	9%	0%	0%	0%	0%	6%	0%
671 Holding Offices	0%	0%	0%	0%	0%	12%	0%	0%	0%
737 Computer and Data Processing Ser.	13%	16%	5%	4%	16%	2%	3%	2%	4%
738 Miscellaneous Business Services	0%	1%	0%	0%	0%	0%	0%	1%	6%
762 Electrical Repair Shops	0%	0%	0%	0%	0%	0%	0%	8%	6%
872 Accounting, Auditing, Bookkeeping	3%	0%	0%	0%	9%	0%	0%	0%	0%
873 Research and Testing Services	5%	6%	8%	0%	0%	0%	1%	1%	7%
874 Management and Public Relations	1%	4%	0%	32%	7%	0%	10%	1%	1%
All Other Industries	2%	3%	1%	0%	2%	1%	0%	6%	4%
Total Jobs	76,377	169,072	13,130	8,152	4,309	66,350	18,177	69,150	67,585
Total Establishments	639	605	124	35	59	103	43	528	494
Number of Different 4-Digit SICs	57	69	23	11	17	26	12	67	47

Source: National Establishment Time-Series (NETS Database, 1990-2003
Walls & Associates

Table 3: Functional Job Composition in 2002: High-Tech Services Firms

	Microsoft	Oracle	Novell	Peoplesoft	Computer Associates	BMC Software	Siebel Systems	Intuit
(SIC) Codes	%Jobs	% Jobs	% Jobs	% Jobs	% Jobs	% Jobs	% Jobs	% Jobs
349 Misc. Fabricated Metal Products	0%	0%	0%	0%	0%	0%	0%	0%
357 Computer and Office Equipment	1%	0%	0%	0%	0%	0%	0%	0%
366 Communications Equipment	0%	0%	0%	0%	0%	0%	0%	0%
367 Electronic Components and Access.	1%	0%	0%	0%	0%	0%	0%	0%
369 Misc. Electrical Equipment, Supp.	0%	1%	11%	0%	0%	6%	1%	0%
382 Measuring and Controlling Devices	0%	0%	0%	0%	0%	0%	0%	0%
384 Medical Instruments and Supplies	0%	0%	0%	0%	0%	0%	0%	0%
422 Public Warehousing and Storage	0%	0%	0%	0%	0%	0%	0%	0%
481 Telephone Communication	7%	0%	1%	0%	0%	0%	0%	0%
504 Professional and Comm. Equipment	8%	3%	2%	0%	1%	6%	0%	2%
506 Electrical Goods	0%	0%	0%	0%	0%	0%	0%	0%
573 Radio, Television, Computer Stores	0%	3%	0%	0%	1%	0%	0%	0%
596 Nonstore Retailers	0%	0%	0%	0%	0%	0%	0%	0%
599 Retail Stores, NEC	0%	0%	0%	0%	0%	0%	0%	0%
671 Holding Offices	0%	0%	0%	0%	0%	0%	0%	0%
737 Computer and Data Processing Ser.	80%	85%	84%	100%	93%	87%	98%	78%
738 Miscellaneous Business Services	0%	0%	0%	0%	0%	0%	0%	0%
762 Electrical Repair Shops	0%	0%	0%	0%	0%	0%	0%	0%
872 Accounting, Auditing, Bookkeeping	0%	0%	0%	0%	0%	0%	0%	15%
873 Research and Testing Services	0%	0%	0%	0%	0%	0%	0%	0%
874 Management and Public Relations	1%	7%	1%	0%	1%	0%	1%	5%
All Other Industries	3%	0%	1%	0%	4%	0%	0%	1%
Total Jobs	12,095	16,152	9,161	10,369	10,297	6,643	5,911	4,392
Total Establishments	90	149	71	51	134	45	67	49
Number of Different 4-Digit SICs	21	18	15	6	12	8	3	10

Source: National Establishment Time-Series (NETS Database, 1990-2003
Walls & Associates

Table 4: Computer High Technology Jobs, 1989-2002: Top 25 MSAs in Selected Years (The Role of Market Dynamics)

MSA	1989	Rank	1995	Rank	2002	Rank
San Jose, CA	288,791	1	288,791	1	332,282	1
Washington, DC-MD-VA-WV	190,050	2	222,791	3	281,318	2
Chicago, IL	185,548	3	224,355	2	197,989	4
Los Angeles-Long Beach, CA	184,406	4	162,635	5	170,283	6
Boston, MA-NH	184,080	5	160,033	6	173,880	5
Dallas, TX	158,603	6	190,314	4	214,148	3
Orange County, CA	123,659	7	98,234	10	121,354	9
Minneapolis-St Paul, MN-WI	110,689	8	83,720	12	101,235	14
Philadelphia, PA-NJ	102,783	9	108,486	8	116,759	11
New York, NY	96,367	10	98,479	9	131,110	8
Phoenix-Mesa, AZ	89,833	11	148,901	7	105,174	12
Atlanta, GA	86,209	12	93,952	11	147,960	7
Nassau-Suffolk, NY	69,845	13	66,432	16	57,566	23
San Diego, CA	67,598	14	60,891	19	77,048	19
Seattle-Bellevue-Everett, WA	63,285	15	64,571	17	74,939	20
Raleigh-Durham-Chapel Hill, NC	55,687	16	70,545	13	72,291	21
Houston, TX	54,650	17	68,894	14	80,655	18
Portland-Vancouver, OR-WA	54,239	18	55,470	21	83,716	16
San Francisco, CA	54,021	19	56,689	20	102,790	13
Detroit, MI	52,237	20	68,396	15	82,984	17
Dutchess County, NY	49,430	21	53,184	23	21,542	58
Baltimore, MD	49,190	22	44,502	26	52,454	25
Colorado Springs, CO	48,494	23	43,496	27	23,354	52
Oakland, CA	47,626	24	52,694	24	117,074	10
Newark, NJ	45,894	25	45,497	25	52,613	24
Denver, CO	43,481	26	55,444	22	92,459	15
Austin-San Marcos, TX	42,884	27	62,096	18	67,151	22
St Louis, MO-IL	39,502	29	42,006	28	43,265	28
Top 25 MSAs:						
Jobs	2,524,215		2,661,494		3,107,332	
% of Total Sample	55.0%		55.7%		56.5%	

Source: National Establishment Time-Series (NETS Database, 1989-2002
Walls & Associates

Table 5: Firm-Level Geographic Measures 1989-2002

Variable	Mean	Standard Deviation	Minimum	Maximum
Inner Network Interactivity (000)	2.53	16.55	0.00	246.78
Outer Network Interactivity (000)	1900.76	14,996.47	0.00	414,737.05
Total Establishment-level Access to Employees in Computer-Manufacturing SICs	96,326	590,411	423	2,737,325
Total Establishment-level Access to Employees in Computer-Services SICs	473,402	347,343	683	1,550,146
N = 2168				

Table 6: Firm-Level Output and Factor Inputs, 1989-2002 (Deflated to 1989 $)

Variable	Mean	Standard Deviation	Minimum	Maximum
Numbers of Employees (000)	12.018	32.27	0.512	383.22
Number of Establishments	73.43	112.12	1.00	637.00
Capital Stock ($000,000)	556.25	1936.24	.067	26,078.94
Research and Development Stock ($000,000)	850.45	2,805.62	0.00	42,239.69
Total Sales ($000,000)	2,707.58	7,673.93	.22	75,780.00
Other Purchased Inputs ($000,000)	1,509.71	4,250.59	.25	42,239.69
N = 2168				

Table 7: Production Function Estimates: Cobb-Douglas and Translog Estimates (Specification Includes Fixed Effects, Annual Controls, White Standard Errors, 177 Firms 1989-2002 in an Unbalanced Panel, N=2168)

	Cobb Douglas		Translog	
Coefficient	Coefficient Estimate	Standard Error	Coefficient Estimate	Standard Error
Intercept	-0.043***	0.013	-0.037***	0.013
Ln(Number of Employees)	0.143***	0.014	0.162***	0.015
ln(Other Purchased Inputs)	0.551***	0.012	0.561***	0.013
ln(R&D Stock)	0.0002	0.004	0.0001	0.41
ln(Capital)	0.224***	0.012	0.213***	0.013
ln(Inner Network Interactions)	0.002	0.003	0.004**	0.002
ln(Outer Network Interactions)	0.005**	0.002	-0.001	0.002
ln(Total MSA Employment in Computer Manufacturing SIC)	-0.093***	0.016	-0.093***	0.016
ln(Total MSA Employment in Computer Services SIC)	0.101***	0.016	0.102***	0.016
Sq. of ln(Number of Employees)			0.033***	0.007
Sq of ln(Other Purchased Inputs)			0.064***	0.013
Sq of ln(R&D Stock)			0.00009	0.0007
sq of ln(Capital Stock)			-0.031**	0.014
ln(employment)*ln(Capital Stock)			0.047	0.013
ln(employment)*ln(Other Inputs)			-0.095***	0.016
ln(employment)*ln(R&D Stock)			0.009	0.006
ln(Capital Stock)*ln(R&D Stock)			-0.010**	0.004
ln(Capital Stock)*ln(Other Inputs)			0.017*	0.009
ln(R&D Stock)*ln(Other Inputs)			-0.002	0.004
	R^2	.928	R^2	.93

$***$ Statistically significant at better than the .01
$**$ Statistically significant at better than the .05
$*$ Statistically significant at better than the .10

Table 8: Production Function REML Estimates: Cobb-Douglas with random factor coefficients (N=2168)

	REML Cobb Douglas	
Variable Factor Input Estimates	Coefficient Estimate	Standard Error
Coefficient on Intercept)		
Intercept	-6.020***	3.019
ln(Inner Network Interactions)	-0.093***	0.049
ln(Outer Network Interactions)	-0.098**	0.034
ln(Total MSA Employment in Computer Manufacturing SIC)	-0.060	0.304
ln(Total MSA Employment in Computer Services SIC)	0.721**	0.291
Coefficient on Ln(Number of Employees)		
Intercept	-1.673***	0.518
ln(Inner Network Interactions)	-0.027**	0.012
ln(Outer Network Interactions)	-0.019**	0.009
ln(Total MSA Employment in Computer Manufacturing SIC)	0.042	0.078
ln(Total MSA Employment in Computer Services SIC)	0.118***	0.060
Coefficient on Ln(Other Purchased Inputs)		
Intercept	2.162***	0.472
ln(Inner Network Interactions)	0.020**	0.10
ln(Outer Network Interactions)	0.016*	0.007
ln(Total MSA Employment in Computer Manufacturing SIC)	-0.007	0.060
ln(Total MSA Employment in Computer Services SIC)	-0.120**	0.062
Fixed Factor Input Estimates		
ln(R&D Stock)	0.048***	0.012
ln(Capital Stock)	0.128***	0.011
Variable Factor Input Technology	Mean	Standard Deviation
ln(Employment)	0.273	0.319
ln(Other Purchased Inputs)	0.65	0.298

$***$ Statistically significant at better than the .01
$**$ Statistically significant at better than the .05
$*$ Statistically significant at better than the .10

Figure 1: SIC Components of the Computer Cluster: Total U.S. Employment (1989-2002) Source: Employment, Hours and Earnings from the Current Employment Statistics Survey, Bureau of Labor Statistics

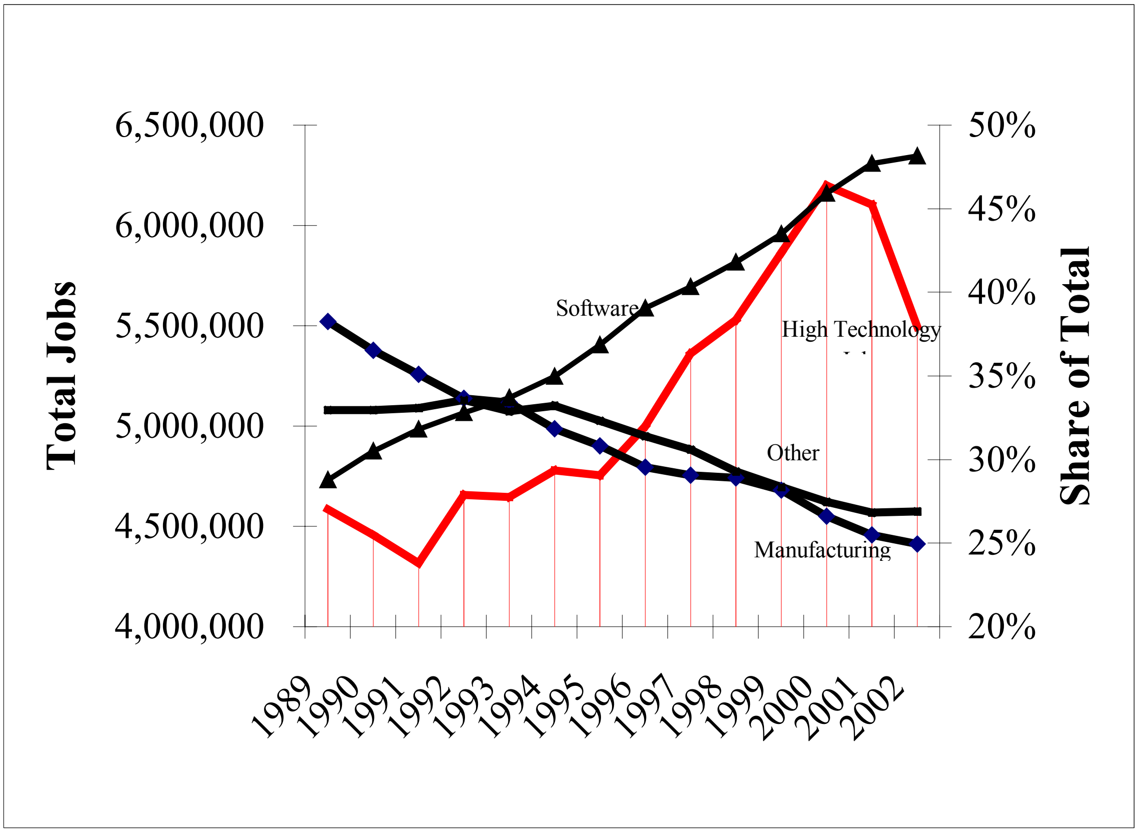


Figure 2: SIC Components of the Computer Cluster: 177 of the Largest Publicly-Traded Firms (1989-2002) (Source: National Establishments Time Series Data)

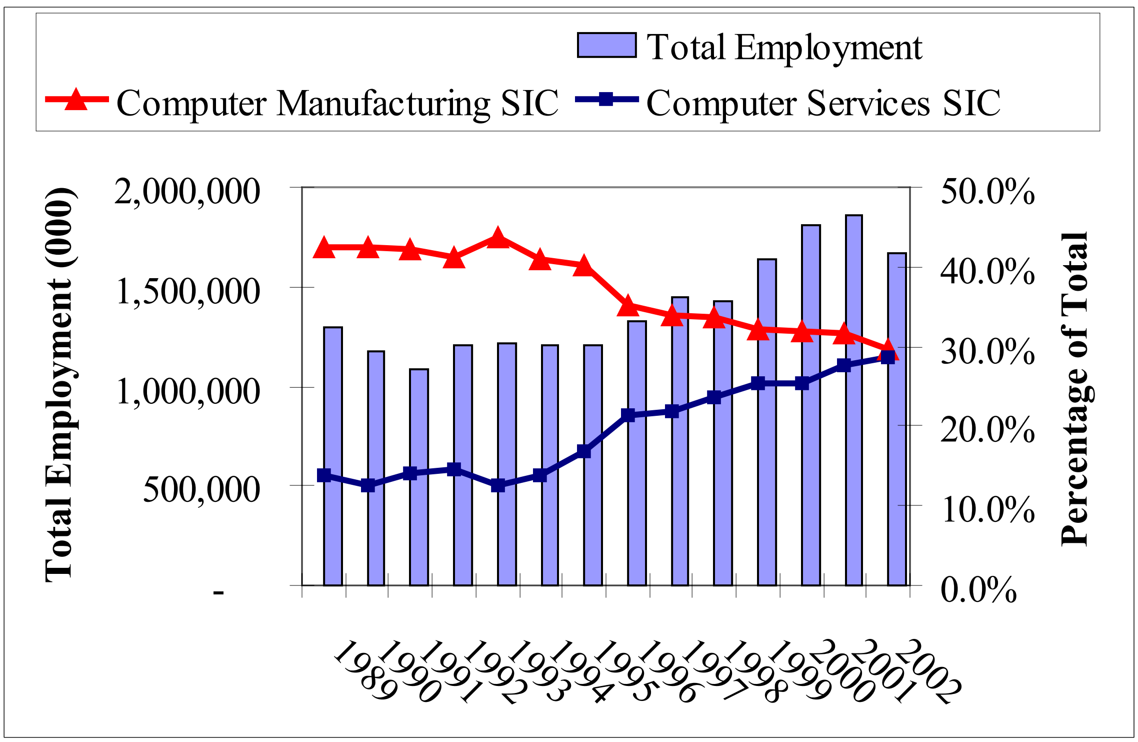


Figure 3: Establishment Locations: Sample of the Largest Publicly-Traded High-Tech Firms in the Computer Cluster (Source: National Establishments Time Series Data)

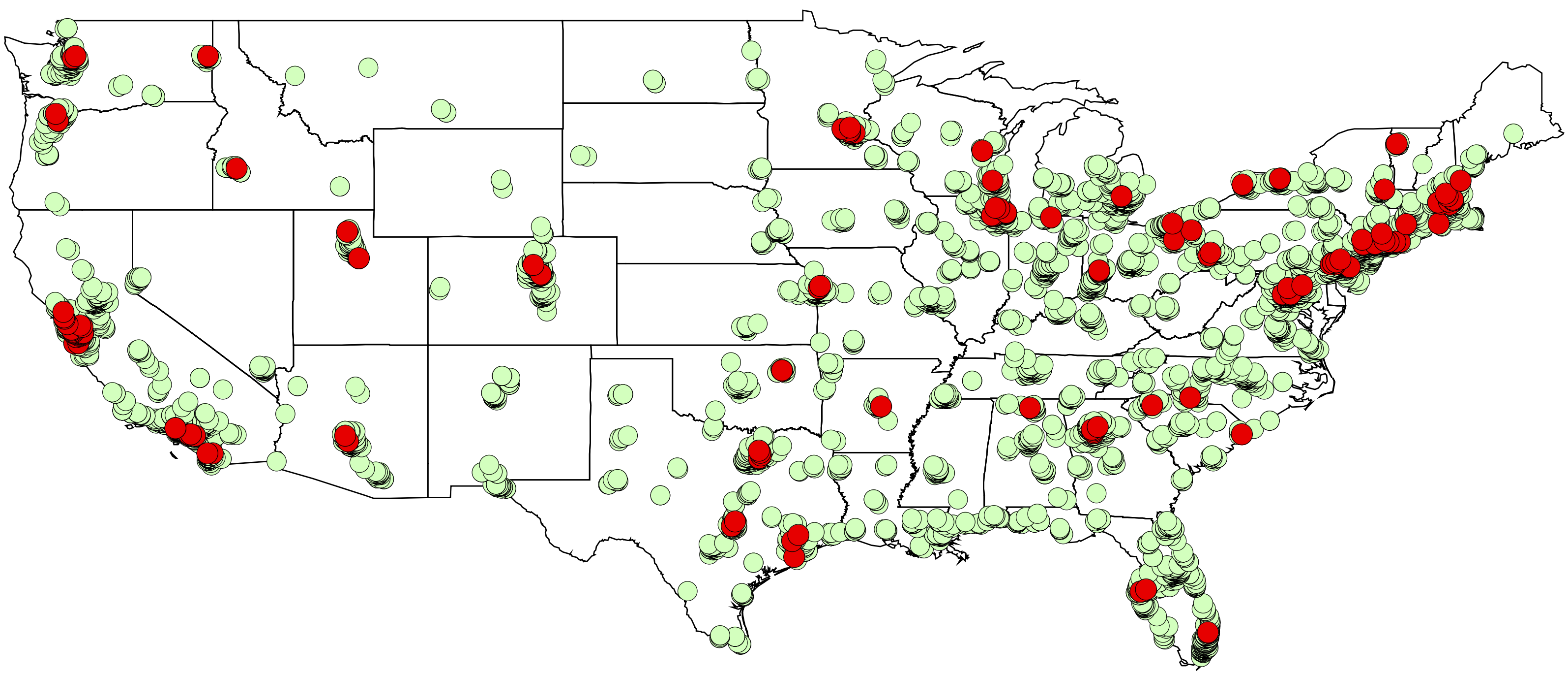


Figure 4: Headquarter Locations: Sample of the Largest Publicly-Traded High-Tech Firms in the Computer Cluster (Source: National Establishments Time Series Data)

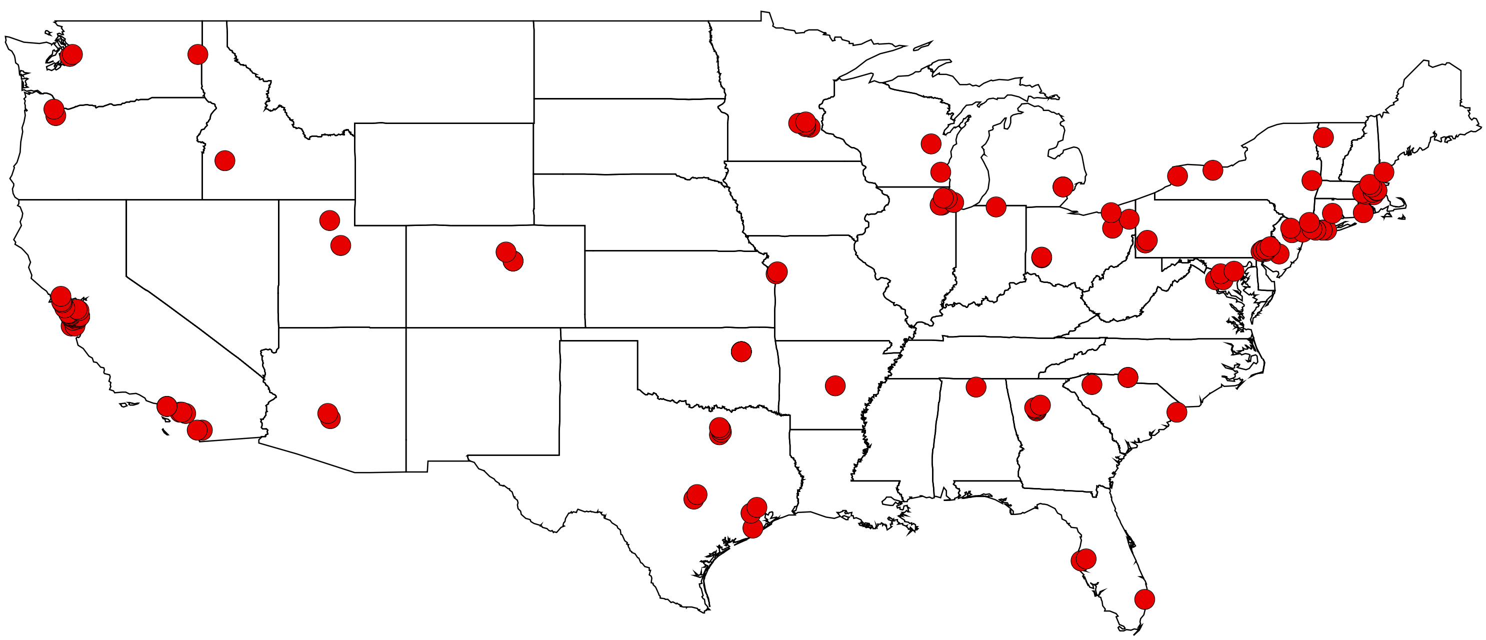


Figure 5: Manufacturing Establishment (NAICS 3341, 3344): Sample of the Largest Publicly-Traded High-Tech Firms in the Computer Cluster (Source: National Establishments Time Series Data)

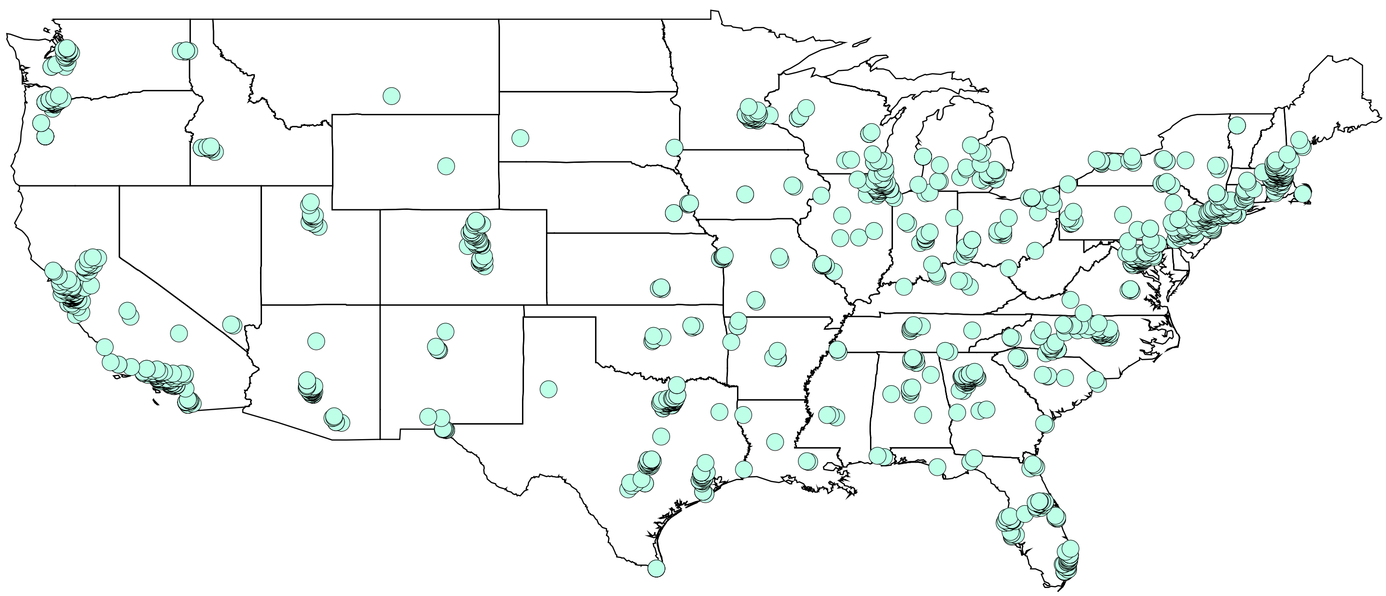


Figure 6: Computer Services Establishments: Sample of the Largest Publicly-Traded High-Tech Firms in the Computer Cluster (Source: National Establishments Time Series Data)

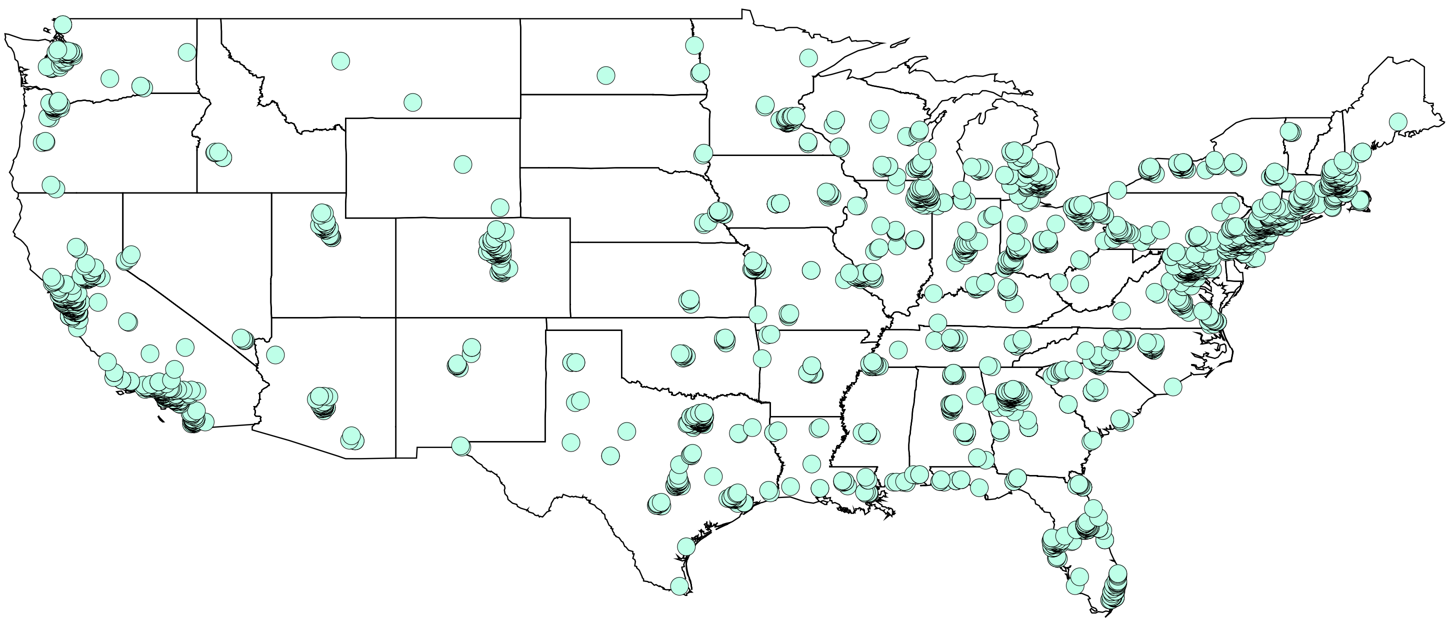


Figure 7: Histogram of the REML Firm-Level Coefficient Estimates on Employment

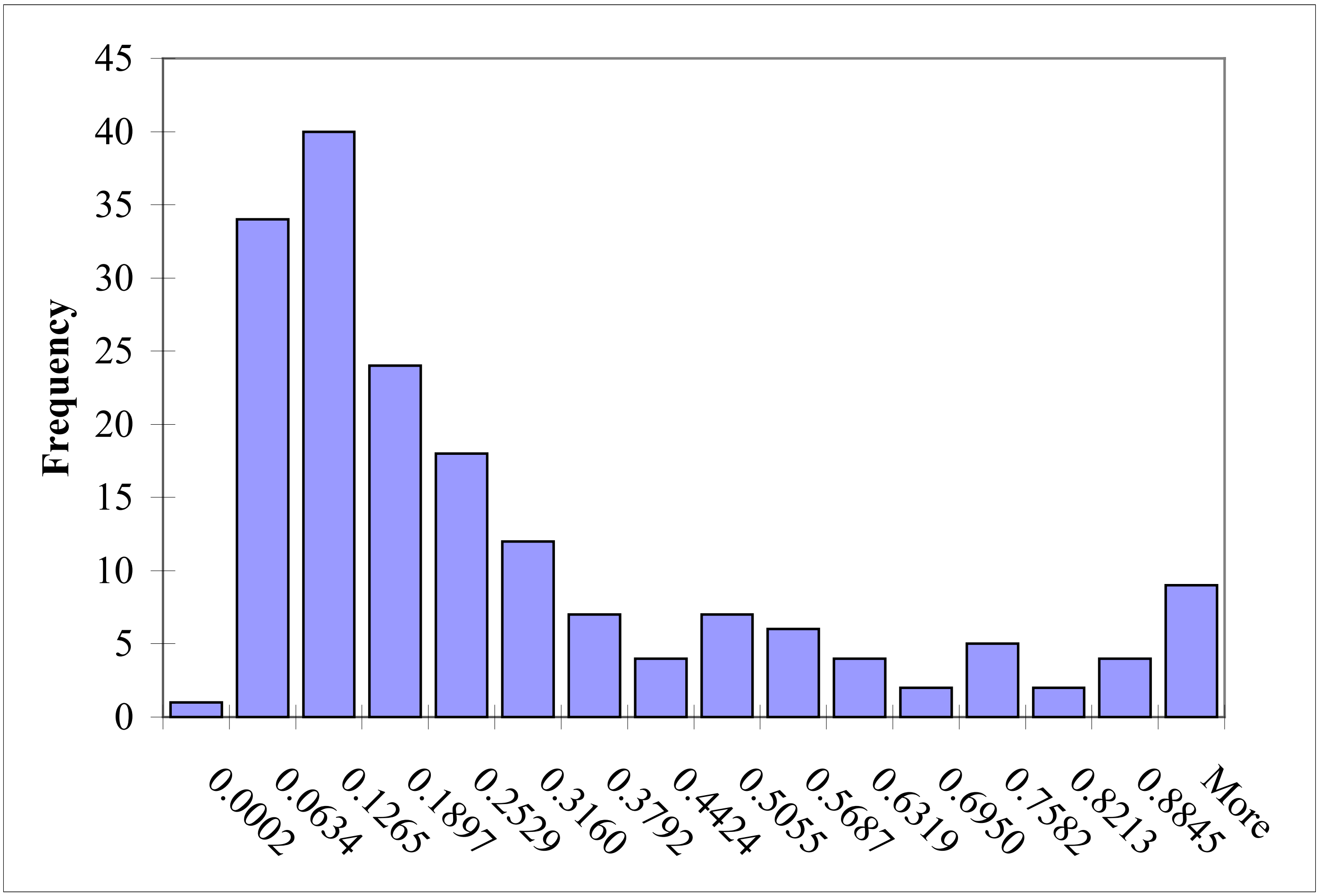


Figure 8: Histogram of the REML Firm-Level Coefficient Estimates on Other Purchased Inputs

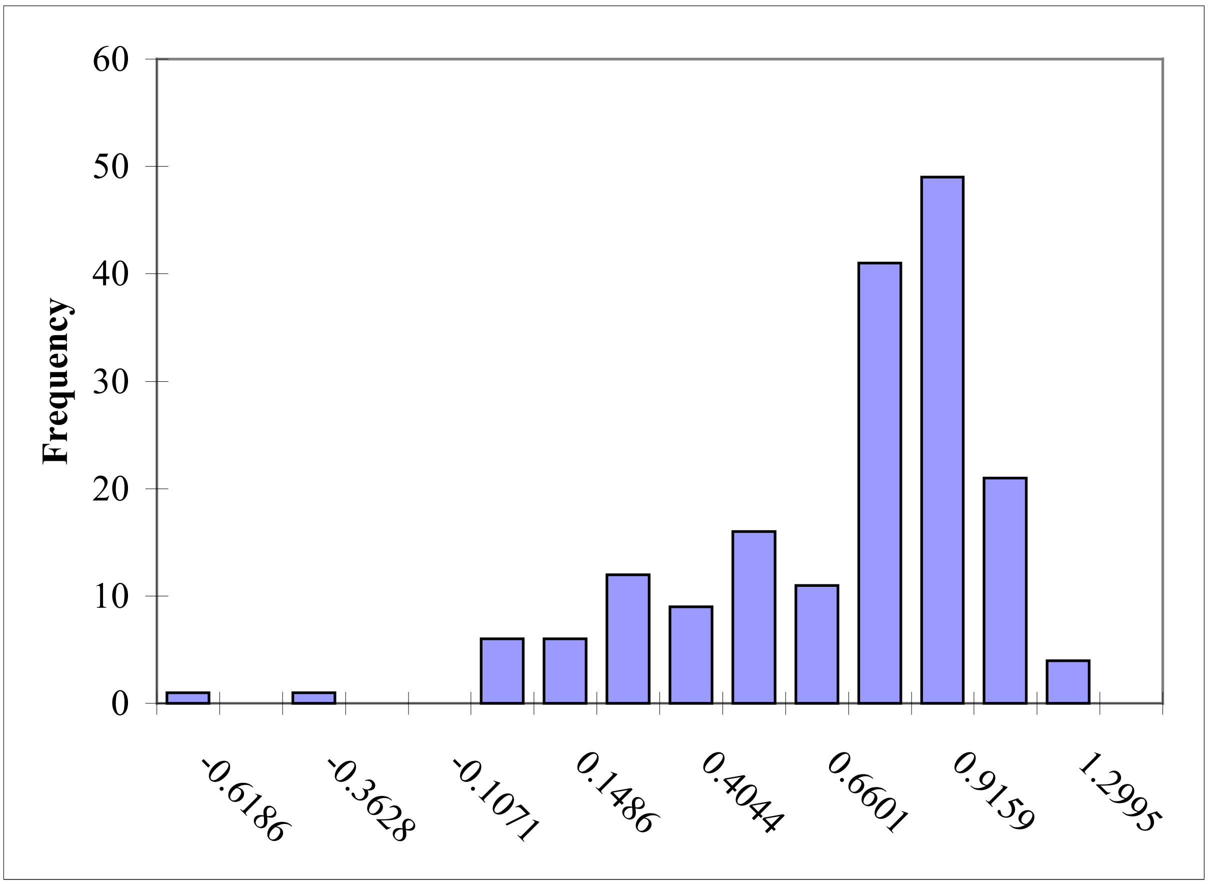


Figure 9: REML Estimates of the Firm-Level Coefficients on Employment and Other Purchased Inputs

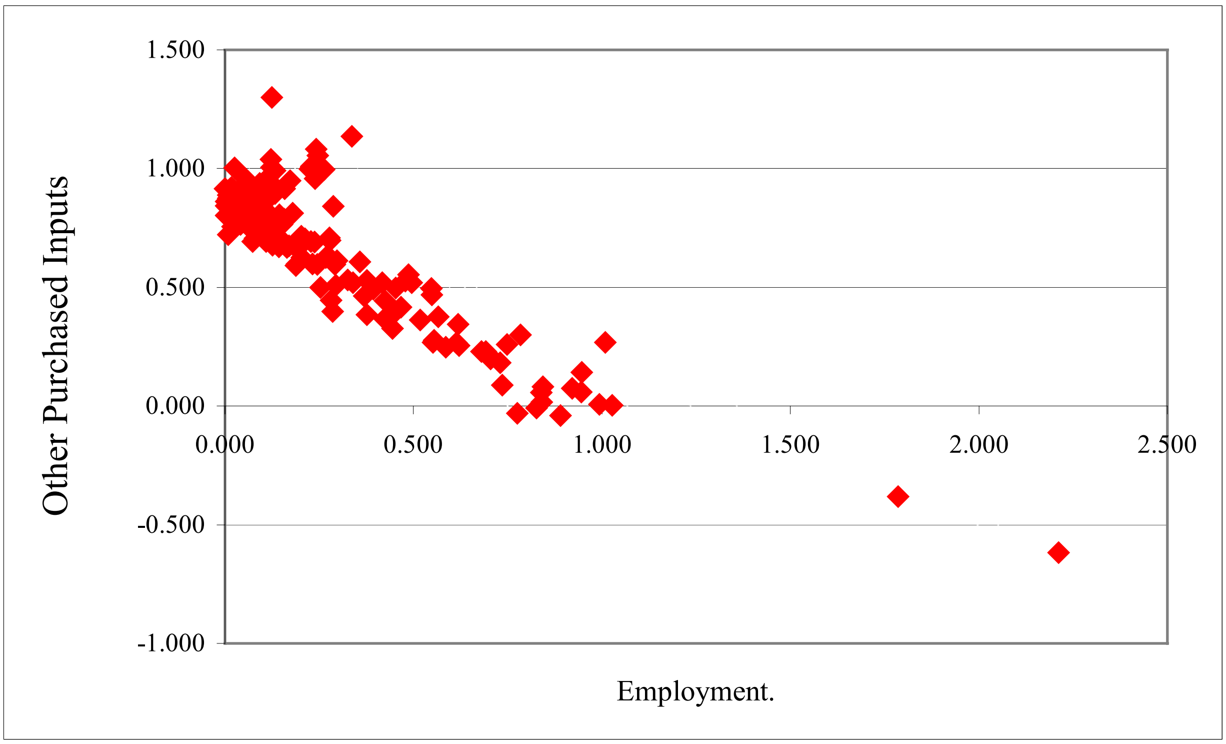
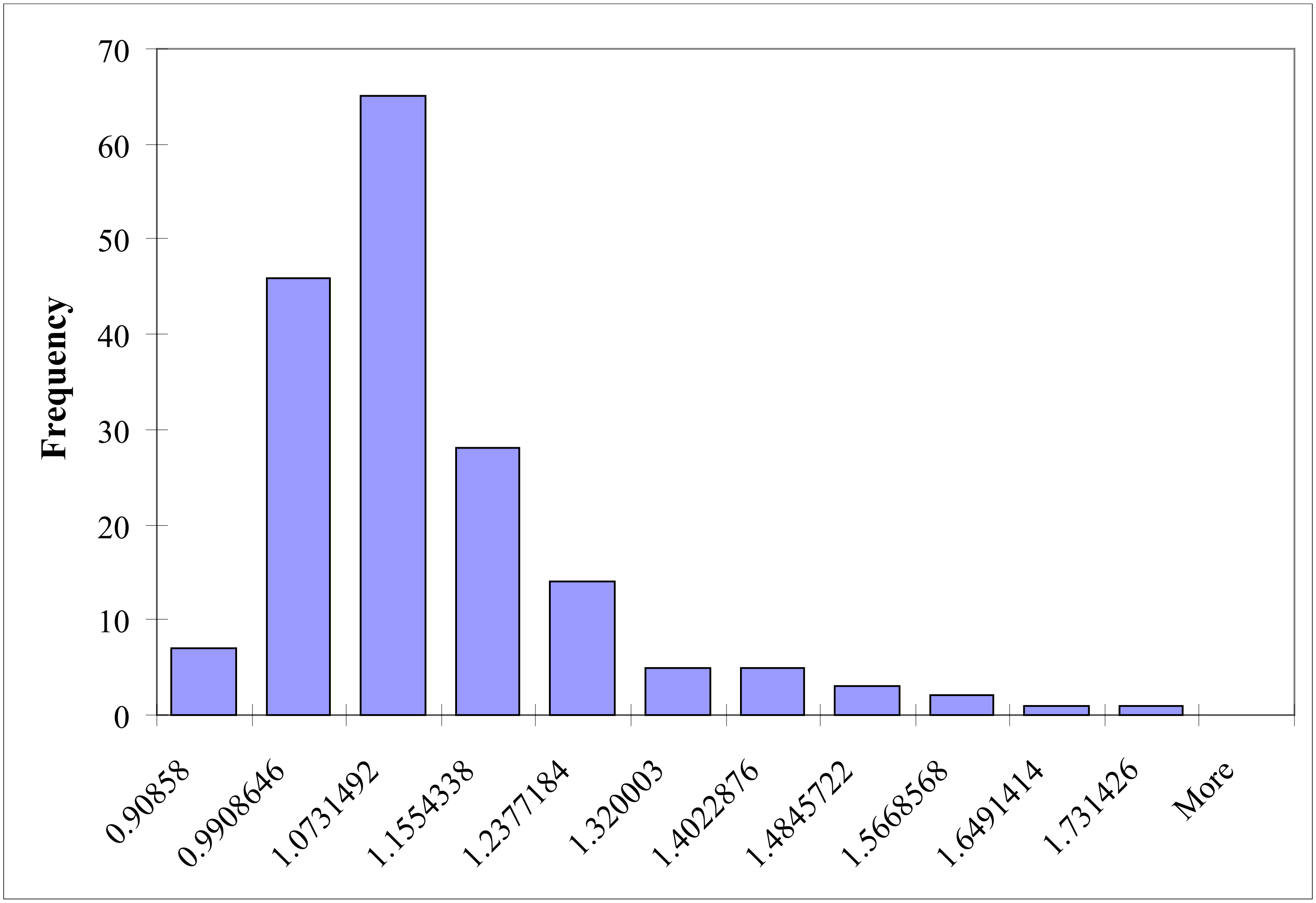


Figure 10: Histogram of the Estimated Firm-Level Returns to Scale



References

Almazan, A., A. de Motta, and S. Titman, 2003, Firm location and the creation and utilization of human capital, Working paper University of Texas, Austin.

Bardhan, A., D. Jaffee, and C. Kroll, 2003, *Globalization of a High-Tech Economy* (Kluwer, New York.).

Combes, P. P., and G. Duranton, 2003, Labour pooling, labour-poaching, and spatial clustering, Working paper London School of Economics.

Dempster, A., D. Rubin, and R. Hall, 1981, Estimation in covariance components models, *Journal of the American Statistical Association* 76, 341–353.

Dumais, G., G. Ellison, and E. Glaeser, 2002, Geographic concentration as a dynamic process, *Review of Economics and Statistics* 84, 193–204.

Duranton, G., and H. Overman, 2004, Testing for localization using micro-geographic data, Working paper London School of Economics.

Duranton, G., and D. Puga, 2001, Nusery cities: Urban diversity, process innovation, and the life cycle of products, *American Economic Review* 91, 1454–1477.

Ellison, E., and E. Glaeser, 1997, Geographic concentration in us manufacturing industries: A dartboard approach, *Journal of Political Economy* 105, 889–927.

Fujita, M., and H. Ogawa, 1982, Multiple equilibria and structural transition of non-monocentric urban configurations, *Regional Science and Urban Economics* 161–196.

Fujita, M., and J. F. Thisse, 2002, *Economics of Agglomeration: Cities, industrial location, and regional growth* (Cambridge University Press, London.).

Fuss, M., D. McFadden, and Y. Mundlak, 1978, A survey of functional forms in the economic analysis of production, in M. Fuss, and D. McFadden, eds., *Production economics: A dual approach to theory and applications* (North Holland, Amsterdam.).

Glaeser, E., H. Kallal, J. Schienkman, and A. Shleifer, 1992, Growth in cities, *Journal of Political Economy* 100, 1126–1152.

Griliches, Z., and J. Mairesse, 1995, Production functions: the search for identification, Technical Report 507, Cambridge, MA.

Hall, B. H., and J. Mairesse, 1995, Exploring the relationship between research and development and productivity in french manufacturing firms, *Journal of Econometrics* 65, 263–293.

Helsley, R., and W. Strange, 1990, Agglomeration economies and matching in a system of cities, *Regional Science and Urban Economics* 20, 189–212.

Henderson, J. V., 2003, Marshall's scale economies, *Journal of Urban Economics* 53, 1–28.

Hsiao, C., 1986, *Analysis of panel data* (Cambridge University Press, Cambridge.).

Jaffee, A., M. Trajtenberg, and R. Henderson, 1993, Geographic localization of knowledge spillovers evidenced by patent citations, *Quarterly Journal of Economics* 108, 577–598.

Krugman, P., 1991, *Geography and Trade* (Cambridge University Press, Cambridge.).

Laird, N., and J. Ware, 1982, Random effects models for longititudinal data, *Biometrics* 38, 963–974.

Marschak, J., and W. Andrews, 1944, Random simultaneous equations and the theory of production, *Econometrica* 12, 143–205.

Marshall, A., 1890, *Principles of economics* (Macmillan, New York.).

Moretti, E., 2004, Worker's education, spillovers, and productivity: Evidence from plant-level production functions, Forthcoming in American Economic Review.

Mundlak, Y., 1978, On the pooling of time series and cross section data, *Econometrica* 46, 69–85.

Mundlak, Y., 1988, Endogenous technology and the measurement of productivity, in S. Capalbo, and J. Antle, eds., *Agricultural productivity: Measurement and explanation* (Resources for the Future, Washington, D.C..).

Mundlak, Y., and R. Hellinghausen, 1982, The intercountry agricultural production function: Another view, *American Journal of Agricultural Economics* 64, 664–672.

Rosenthal, S., and W. Strange, 2001, The determinants of agglomeration, *Journal of Urban Economics* 50, 191–229.

Rotemberg, J., and G. Saloner, 2000, Competition and human capital accumulation: A theory of interregional specialization and trade, *Regional Science and Urban Economics* 30, 373–404.

Schoar, A., 2003, Effects of corporate diversification on productivity, forthcoming in Journal of Finance.

A1 National Establishment Time-Series (NETS) Data

The NETS database is constructed from annual snapshots of the Dun's Marketing Information (DMI) files identifying which establishments were active in January of each year. NETS also includes Dun's archival Credit Rating files that provide annual raw establishment data. The establishment-level data includes: the business name, address and contact information (including officer, title, phone number, FIPS codes and longitude and latitude); headquarters linkages (including the D-U-N-S Number of the topmost domestic firm in a "Family Tree" of companies, as well as the parent company and headquarters, the location of the establishment within the corporate hierarchy, and whether the ownership has changed 1990-present; years when business was active ("1989" is earliest year in Database and, currently, "2003" is latest year) and year business started; industry classification (primary 8-digit SIC and up to five secondary SICs; whether the primary 3-digit SIC changed 1990-present); type of establishment (Single location, headquarters, or branch; public or private; and legal status: proprietorship, partnership, corporation or non-profit); employment at location and job growth relative to peers (3-digit SIC); estimated annual sales at the establishment and its sales growth relative to peers; Dun and Bradstreet credit ratings and PayDex Scores (January, minimum and maximum for previous year).

An essential component of the NETS data is the Dun and Bradstreet D-U-N-S Number. Any business location with unique, separate and distinct operations is eligible for a D-U-N-S Number. Business entities include proprietorships; partnerships; corporations and government bodies; self-employed individuals such as doctors, lawyers and contractors; branches and divisions of companies, including lock boxes used exclusively for remittances; and other single locations, parents, headquarters and subsidiaries.