ACCESS TO WORKERS OR EMPLOYERS?

AN INTRA-URBAN ANALYSIS OF PLANT LOCATION DECISIONS

by

Mark J. Kutzbach *
U.S. Bureau of the Census

CES 10-21 September, 2010

The research in this paper was undertaken while the author was at the University of California, Irvine.

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see **www.ces.census.gov** or contact Cheryl Grim, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies 2K130B, 4600 Silver Hill Road, Washington, DC 20233, **CES.Papers.List@census.gov**.

Abstract

This analysis attributes economies of agglomeration to either labor market pooling or employer-based productivity spillovers by distinguishing the effect of access to workers, measured by place-of-residence, from the effect of access to employers. New establishment location choices serve as a measure of productivity advantages, while census tract level data on access to same-industry employment, other-industry employment, and specialized workers, as well as metropolitan area fixed effects, measure sources of agglomeration and other locational characteristics. The four industries included are selected so that each relies on a workforce with a specialized occupation that is identifiable by place-of-residence, and that productivity and cost advantages are the primary drivers of location choice. The results show that both access to specialized workers and access to same-industry employers contribute to economies of agglomeration at an intra-urban spatial scale, and that the magnitude of the worker effect is large relative to employer-based productivity spillovers.

Keywords: Economies of Agglomeration; Labor Market Pooling; Commuting

* This research was conducted while the author was at the University of California, Irvine, Department of Economics. Present affiliation: U.S. Census Bureau, Center for Economic Studies. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. The research in this paper does not use any confidential Census Bureau information. I am grateful to Jan Brueckner, David Neumark, and Kenneth Small, for their advice and guidance, as well as to David Brownstone, Mike McNally, and seminar participants at UCI. I am also thankful to the Department of Economics and School of Social Sciences at UCI, the University of California Transportation Center (UCTC), and the Kassouf Foundation for their generous financial support; to the Southern California Association of Governments and San Diego Association of Governments for use of their travel time data; to the Department of Economics for the use of its California extract from the National Establishment Time Series (NETS); to the Public Policy Institute of California (PPIC) for the use of geocoding data for NETS; and to the Institute for Transportation Studies at UCI for use of computing facilities and travel support. This paper was previously published as a chapter of my dissertation at UCI (2009).

1 Introduction

A new firm determining where to locate within an urban area must consider a multitude of locational factors including facility costs, customers, and regulations. In addition, as is explained in theoretical analyses of agglomeration, a firm locating near other, same-industry firms may receive knowledge and technology spillovers that enhance productivity (Helsley, 1990), and a firm locating in a thick labor market may benefit from lower search costs and higher quality job matches that also result in higher productivity (Helsley and Strange, 1990). Empirical studies have confirmed the importance of proximity to other businesses in both analyses of the relative spatial concentration of industries[1] and in analyses estimating the effect of spatial proximity on productivity (or on proxies thereof).[2] The first set of analyses, studying relative industry concentration, have found that industries relying on innovation or specialized workers tend to be more spatially concentrated, or agglomerated. However, the second set of analyses, on productivity effects, have not disentangled the effect of proximity to business clusters from the effect of proximity to thick labor markets. The contribution of this paper is to use detailed geographic data on both employer location and worker place-of-residence to measure the separate agglomeration benefits of access to employers and access to workers, all at an intra-urban spatial scale. In doing so, the paper examines the determinants of productivity differences within cities to explain why firms may be attracted to specific areas of cities.

Because residential areas are dispersed from business districts within metropolitan areas, the relative access to employers and access to workers may vary across a city. This dispersion is evident from studies of urban geography, which find job-housing imbalance

[1] For example, Rosenthal and Strange (2001) and Ellison et al. (2007). See further discussion of these and other empirical studies on the sources of economies of agglomeration in Section 2.

[2] Studies finding productivity spillovers at the intra-urban spatial scale include: Wallsten (2001), Rosenthal and Strange (2003), Van Soest et al. (2006), Ahronson et al. (2007), and Graham (2007).

within cities, with some areas having more jobs than resident workers and others having more resident workers than jobs.[3] There are several reasons why workers might not reside in the immediate area of their job. First, because moving is costly, workers may seek to live near the "average job," rather than any specific employer, based on the assumption that they may switch jobs in the future (Crane, 1996; Yang, 2008). Second, two worker households are constrained from locating optimally for both workers' employment locations (Kim, 1995; Freedman and Kern, 1997). Third, workers may locate based on proximity to non-work destinations and local amenities.[4] Thus, total employment at a location only partially describes a location's access to workers who might commute there. Theoretical analyses suggest that employers must compensate workers for longer commutes (Mills, 1972), and empirical studies find that the urban wage gradient compensates for commuting cost, even when controlling for worker characteristics (Timothy and Wheaton, 2001). Therefore, by choosing locations more accessible to workers, businesses may increase their exposure to the productivity benefits of labor market pooling and reduce labor costs.

The present analysis estimates the overall effects of spatial concentration of employers and workers on new establishment employment for a set of four industries. The selected industries have a national distribution scale, so location choices are not based directly on sales to consumers. For each industry, agglomeration benefits, measured at the census tract level, are regressed on measures of access to employers and access to workers. The dependent variable, total new establishment employment, represents the profit maximizing location choice of businesses and proxies for the productivity benefits of

[3] Anas et al. (1998) review the literature on urban spatial structure

[4] All three of these factors contribute not only to job-housing imbalance, but also to "excess commuting," where total urban commuting is greater than the minimum commuting if workers located optimally based on their places of work. See Ma and Banister (2006) for a review of the literature on excess commuting. See Giuliano and Small (1993) for a study using data job-housing imbalance for the Los Angeles region.

agglomeration. The independent variables include measures of access to: same industry employment located in and around the census tract, indicating a potential for productivity spillovers; other industry employment, indicating overall urban density; and workers specialized in an occupation crucial to the industry, indicating labor market thickness. Thus, although the present analysis does not specify the precise mechanism of economies of agglomeration, such as the number of patent citations (Jaffe et al., 1993), it does disentangle the overall effect of access to workers from the overall effect of access to employers. In addition to these variables, Metropolitan Statistical Area (MSA) level fixed effects control for city-wide agglomeration benefits that do not vary at the intra-urban scale. The analysis is conducted with data for California and uses National Establishment Time Series (NETS) data on employers and Decennial Census demographic data on workers.

For three of the four industries included, this analysis finds that thick labor markets contribute to economies of agglomeration. One complication at the intra-urban scale is the positive correlation of access to workers and access to employers, even within industries. Nevertheless, regressions including both access to same-industry employment and access to workers, while controlling for access to other-industry employment, find that both have positive and significant effects on new establishment employment.[5] Moreover, the magnitude of the labor market pooling effect is large compared to the access to employers effect. In addition, regressions using travel time to measure access, rather than distance, find an even greater relative effect of access to workers, underscoring the importance of commuting access for labor market pooling to take place. Overall, these results suggest that agglomeration benefits depend on access to both employers

[5]Additional regression specifications including either the access to employment variables alone or access to workers variables alone, both find a significant and positive effect of access to either source on new establishment employment. However, because of the correlation of the access variables, such regressions do not establish the source of agglomeration benefits.

and workers at an intra-urban spatial scale.

2 Background

The idea underlying the empirical model in this paper is that businesses derive distinct benefits, through different mechanisms, from locating among industry clusters and in thick labor markets. Marshall (1920) outlines many of these mechanisms and Chinitz (1961) proposes a web of productivity enhancing links within and between industries and also to workers. Recent empirical work has found evidence that these links influence productivity.[6]

Benefits from locating in business clusters, hereafter referred to as "benefits of access to employers (or employment)," include knowledge spillovers, concentrated human capital, and economies of scale in input production.[7] These benefits require proximity to an actual location of production, or place of work. A first benefit, spillovers of specialized knowledge, or "the secrets of the trade," is not bought or sold; rather, it may occur through regular formal and informal interaction of nearby establishments. Early empirical evidence of spillovers was provided by Jaffe et al. (1993), who finds that patent citations are highly spatially concentrated. Similarly, Audretsch and Feldman (1996) find that innovative activity is more spatially concentrated in knowledge-oriented industries. Second, the concentration of human capital and skills in cities makes workers more productive, as is shown by higher wages. As evidence, Rosenthal and Strange (2008) estimate the effect of the density of high skill employment on wages using place-of-work Census data. Accounting for the endogeneity of employment density by using instrumental variables for potential density (using geological data that determines potential density), they

[6]See Rosenthal and Strange (2004) and Duranton and Puga (2004) for surveys of the literature on the empirical and theoretical microfoundations of economies of agglomeration.

[7]Benefits from access to same-industry employers are also known as economies of localization, while benefits of access to other-industry employers are also known as economies of urbanization.

find that proximity to high skill workers increases wages.[8] Third, clustered businesses allow for input providers to attain economies of scale and lower input costs. For example, Holmes (1999) shows that industries that are more vertically disintegrated, and have greater purchased input intensity, tend to be more spatially concentrated.

Locating in thick labor markets offers the distinct benefit of labor market pooling, also referred to in this paper as a "benefit of access to workers." This benefit requires proximity to workers themselves, who commute from their places of residence. A large employment cluster effectively pools its labor force, reducing search costs for both employers and workers, and allowing for more job switching. For example, Fallick et al. (2006) and Freedman (2008) find evidence of especially high worker mobility in California's spatially concentrated Silicon Valley computer industry. In a more comprehensive analysis, Bleakley and Lin (2007) find that in more dense urban areas, young workers switch jobs more rapidly and older workers switch less rapidly, possibly indicating a matching process whereby young workers rapidly identify superior industry and occupation matches and then stick with them. This assortative matching results in higher quality long term matches, improving productivity. Analyzing the mechanism of labor market pooling directly, Andersson et al. (2007) estimate production functions using the employee/employer matched microdata. They show that complementarities between worker and firm quality increase productivity, and that assortative matching increases with urban population density. A related benefit of labor market pooling is reduced risk: for firms, of being without a critical type of skilled labor; and for workers, of being without employment. Overman and Puga (2009), using UK data on manufacturing establishments, find that industry sectors whose establishments experience more idiosyncratic

[8] One means by which density may increase knowledge spillovers is through improved workplace communication (Charlot and Duranton, 2002). As evidence of the importance of communication for productivity, Bacolod et al. (2008) find that cities concentrate employment of workers with high cognitive and people skills, generating higher wages for those with such skills.

volatility are more spatially concentrated. They control for other industry characteristics that are predictive of concentration and conclude that businesses susceptible to volatility use labor market pooling to smooth out shocks.

To examine the relative importance of the mechanisms discussed above, a number of inter-industry studies compare the spatial concentration of industries and find an important role for labor market pooling as a reason for agglomeration, but do not examine its effect on productivity directly. These inter-industry studies follow two empirical methods. One approach (Audretsch and Feldman, 1996; Rosenthal and Strange, 2001) is to estimate the effect of industry characteristics on the relative geographic concentration of industries. These studies find that industries with highly skilled and specialized labor forces are more likely be be geographically concentrated at the ZIP code, county, and state levels. Another approach (Dumais et al., 1997; Ellison et al., 2007; Kolko, 2008) is to estimate the effect of industry pair complementarities on co-agglomeration of production among those industries. These studies find that industries with related labor forces, as measured by the occupation designations of workers, are relatively more concentrated at the ZIP code, county, and state levels.

Of the studies that do examine the direct effects of agglomeration employment or productivity, all focus on the location of business clusters as a source of benefits, and do not consider the distinct benefits from access to employers and access to workers. City-level studies use aggregate urban employment, by industry, to explain city/industry employment growth or productivity (Glaeser et al., 1992; Henderson et al., 1995). However, a city-level study cannot distinguish among locations within cities, where productivity differences may attenuate rapidly over distance. Using detailed geographic data on employment location, Wallsten (2001), Rosenthal and Strange (2003), Van Soest et al. (2006), and Aharonson et al. (2007) find rapid attenuation and attribute it primarily to the importance of face to face encounters for knowledge spillovers. As in the present

paper, they estimate the effect of local characteristics on the location choice of new establishments, as a proxy for productivity effects. In addition, Rosenthal and Strange (2003) show that intra-urban differences in agglomeration benefits are large relative to inter-urban differences in agglomeration benefits. Graham (2007) uses data on firm output to estimate the productivity effect of access to other employers for a set of industries, also finding that benefits attenuate with distance.

3 Methodology

The empirical approach in this paper is similar to that of Rosenthal and Strange (2003), which tests for economies of agglomeration by examining whether new establishment births in an industry occur near concentrations of existing employment in that industry. The present analysis includes several locational characteristics that may affect the profitability of new establishments, including concentrations of existing, same-industry employment, a thick labor market, and other-industry employment. While a positive effect of same-industry employment on new establishment employment is interpreted as a productivity spillover, a positive effect of a thick labor market would be interpreted as a benefit of labor market pooling.

The primary advantage of focusing on new establishments is that, in choosing a location, they take the distribution of existing establishments and the labor force as given. Thus, locational characteristics may be treated as being exogenous to the spatial distribution of new establishments. In contrast, if employment concentrations are endogenous to the productivity of existing establishments, then an estimate of the effect of employment concentrations on the growth of existing establishments may be upward biased.[9] An additional benefit of focusing on new establishments is that their distribution may be

[9]Rosenthal and Strange (2008) overcome this endogeneity problem in their analysis of the effect of employment density on wages, a measure of productivity, by using instrumental variables for density.

more sensitive to locational characteristics. Whereas fixed relocating costs may dissuade existing establishments from moving to a marginally more productive location, new establishments will have start up costs regardless, and thus may be more flexible in their location choice and more responsive to local productivity differences.

One caveat with examining new establishment employment is that it does not measure whether net employment growth is positive, which also depends on establishment expansions and contractions, deaths, and relocations.[10] In particular, a high volume of births in an area may be offset with an equally high volume of deaths, with young establishments being especially likely to die.[11] While this process of employment reallocation, known as churning, has been found to reallocate employment from less productive firms that exit to more productive firms that enter (although entrants are not necessarily immediately profitable),[12] it is not clear whether a higher rate of churning is necessarily a product of economies of agglomeration.

Nevertheless, the present analysis focuses not on the amount of churning in an urban area, but on the intra-urban spatial allocation of new establishments at the census tract level. As in Rosenthal and Strange (2003), MSA fixed effects are meant to account for fiscal policies, wage rates, and natural advantages, as well as the overall "birth potential" of an urban area due to deaths or downsizing of existing firms (Carlton, 1983). Thus, the assumption here is that while a high degree of churning or birth potential in an urban area may lead to more new establishment births, new establishments will still attempt to locate in an optimal location within the urban area.

Rosenthal and Strange (2003, pp. 378-379) provide a detailed explanation for how

[10] Among these components, establishment births and deaths are responsible for the majority of long run employment reallocation. Neumark et al. (2007), looking at employment reallocation in California, find that in each three year period between 1992 and 2002, the major share of statewide new employment came from new establishments, as opposed to being from the expansion of existing establishments.

[11] Davis et al., 2006

[12] Foster et al., 2008

productivity enhancing locational characteristics affect the profitability of new establishments, and translate into more new establishment births in more productive census tracts. They define new establishment profits as a function of a vector of local characteristics, y, that increase profits by shifting up the production function. Meanwhile, establishment's choose input quantities, such as employment, to maximize profits. With optimal input quantities for a location, a new establishment is born only if it can earn nonnegative profits. Rosenthal and Strange explain that establishments are heterogenous in potential profitability, so the probability of an establishment coming into existence at a location will increase if that location has favorable characteristics.

Suppose there are J census tracts available for a new establishment to locate in. For each census tract, $j = 1, ..., J$, the base year characteristics may be partitioned into two types of variables. A tract's access to employers and access to workers is denoted as a set of variables, y_j, that vary with the tract's intra-urban location. Other characteristics of a tract's metropolitan area are denoted by a set of indicator variables, y_m, for each MSA, $m = 1, ..., M$. Census tract access and metropolitan status shift an establishment's production function, so that some census tracts are more or less productive for a particular industry.

A highly productive census tract has a higher probability of inducing new establishments to locate there, resulting in higher new establishment employment, denoted $NewEmp_j$. The regression equation is specified as

$$NewEmp_j \;=\; \beta_0 + \beta y_j \;+\; \beta_m y_m + \epsilon, \tag{1}$$

where β_0 is a constant term, the parameters in β give the linear effect of the census tract access variables, the set of parameters in β_m give MSA fixed effects, and ϵ is an independent and identically distributed error term. Note that only zero or positive new

establishment employment is observed, obscuring differences in the potential productivity of census tracts with no new establishments. Because this latent productivity is censored, the present analysis, like Rosenthal and Strange (2003), employs a tobit regression model.

While this regression model has the advantage of treating locational characteristics as exogenous, there are some concerns to address. First, the model assumes that the size of new establishments is not related to locational characteristics, so that more births will correspond with more new establishment employment. However, particularly large new establishments may have specific demands for contiguous real estate or traffic capacity. If employment density were negatively correlated with the availability of such locations, a variable that is not available, then the estimated effect of employment density might be biased downwards. Excluding the largest new establishments reduces the impact of this possible bias. Likewise, particularly small new establishments, such as the self-employed and those working from home, may have little demand for labor and may not be sensitive to labor market thickness. As a result particularly small establishments might favor residential census (where the owners live), which could also bias the effect downward. To avoid the extreme ends of new establishment size, the domain of new establishments is restricted to those with less than 500 employees and those with more than one employee.

Second, the model assumes that productivity increases will translate into greater profitability. However, the rent of higher productivity may be shared by establishments with workers and land owners. There is evidence at the firm level linking higher wages to productivity and profits (Haltiwanger et al., 2007) as well as agglomeration in particular (Rosenthal and Strange, 2008). Likewise, if landowners recognize that a specific industry may be especially profitable at a given location, they may increase real estate or rental costs for establishments in that industry, such as at a mature office or industrial park (Rauch, 1993). To the extent that rent sharing in either form is present, it will attenuate

the estimated effects of proximity to a thick labor market and proximity to same-industry employment on new establishment employment. Thus, the productivity effect - proxied for here by new establishment employment - may represent only the share of benefits captured by businesses themselves.

4 Data and Variables

4.1 Industries

The present analysis uses a set of four narrowly defined industries to highlight the importance of location-based productivity differences. This approach, which benefits from detailed industry data, differs from other studies that examine sets of broadly defined industries. For example, Graham (2007) uses industry sectors including: manufacturing, construction, hotels & catering, transportation, storage & communication, real estate, information technology, banking, finance & insurance, business services, and public services. Van Soest et al. (2006) uses the sectors: consumer services, producer services, and manufacturing. When industries are defined so broadly (e.g., manufacturing), the nature of the productivity effect is less clear. Furthermore, by including retail oriented industries, such as banking or consumer services, the density effect may reflect economies of agglomeration based on both production and consumption.[13] In contrast to those studies, Rosenthal and Strange (2003) require that the industries in their analysis distribute their output at a national or international scale, so that the density effect, β, reflects productivity differences rather than sales advantages. They use a set of narrowly defined industries including software, food products, apparel, publishing & printing, fabricated metals, and machinery. Likewise, Wallsten (2001) focuses on a set of high technology manufactur-

[13] Glaeser et al. (2001) examines how the agglomeration of consumption opportunities also contributes to city growth.

ing industries and Aharonson et al. (2007) focus only on the Canadian biotechnology industry.

In addition to the criteria of Rosenthal and Strange (2003), the present analysis requires that place-of-residence data be available on workers specializing in an occupation crucial to each industry. However, because publicly available place-of-residence data is highly aggregated, there is a limited selection of suitable industries. Although the Decennial Census provides tract level demographic data, it only gives 19 unique 2-digit NAICS industry sector classifications and 33 SOC occupation classifications. While industry codes define the type of activity done at a place-of-work, occupations define the kind of work a person does as well as their most important activities and duties. Linking worker groups to industries by occupation is ideal because new establishments may be attracted to workers of a given occupation regardless of the industry they are currently employed in.[14]

Table 1 gives a set of narrowly defined industries meeting the criteria of Rosenthal and Strange (2003) and having public data available on a specialized worker occupation that is crucial to each industry, to measure labor market thickness. The employment shares of workers by occupation in each industry are calculated from Census 2000 Public Use Microdata Sample for California, which, unlike aggregate tabulations, gives both the industry and occupation of individual workers. The software publishing and computer programming industries (hereafter referred to as Software) are especially reliant on workers in mathematical and computer occupations (including programmers); the motion picture and video production industry (Movie), on arts, design, entertainment, sports, and media occupations (including writers, actors, camera operators, etc.); the research and development industry (R & D), on life, physical, and social science occupations; and

[14]Analyses of co-agglomeration of industry pairs (Ellison et al., 2007) find that industries with broad overlap in the occupations of their workers tend to locate together.

precision instruments industries (Instrument), on architecture and engineering occupations (including both specialists and technicians).[15] Employees in other occupations are less unique to each industry, and are thus less likely to drive location choices. For example, all four industries have a similar share of workers in sales and office occupations, such as administrative assistants.

One limitation of this empirical approach is the short list of industries. The criteria of Rosenthal and Strange (2003) exclude industries with a large service or retail component, such as banking, consulting, or legal industries, even though workers in finance, business operations specialist, and legal occupations are listed in Decennial Census tabulations. While much of the literature on economies of agglomeration is focused on manufacturing industries, the highly aggregated occupation class of most workers in those industries, "production" (included with farming and construction as "Other" in Table 1), would make the place of residence of such workers meaningless as a driver of labor market pooling.

Nevertheless, the four industries are highly varied in industry sector (information, manufacturing, and professional service) and thus are representative of a wider set of industries. In addition, these industries are recognized as having regional clusters in California, the area of this study.[16] The present analysis looks not at the decision of businesses in these industries to locate in California, but at precise, intra-urban, location choices.

[15]The industries are defined by NAICS codes as follows: Software includes establishments that design and publish software (511210) or provide custom computer programming services (541511); Movie includes establishments that produce and distribute (not retail businesses) motion pictures and videos (512110); R & D includes establishments specialized in biotechnology, as well as physical, engineering, and life sciences (541710); Instruments include establishments manufacturing measuring and controlling devices for medical, navigational, and other purposes (3345).

[16]See the California Economic Strategy Panel (2006) for a review of regional industry clusters in California.

4.2 Data sources

Measures of new establishment employment for an industry, *EmpNew*, existing employment for the same industry, *EmpSame*, and existing employment for all other industries, *EmpOther*, are calculated from a California extract of the National Establishment Time Series (NETS). NETS data are ideal for this analysis because they provide a detailed spatial measure of employment and follow establishments over time. Walls & Associates (2003) compiles NETS from archived Dun & Bradstreet establishment level data for the years 1990 to 2008. Dun & Bradstreet uses independent data sources (e.g., the Yellow Pages, business registers, and the internet) to compile lists of businesses, including non-profits and the public sector offices. Each year, Dun & Bradstreet hears from or contacts establishments, making over 100 million phone calls, and reports the address, industry code, and number of employees (as well as other data), as of January. Dun & Bradstreet assigns a unique DUNS number (Data Universal Numbering System) to each business to track it from year to year, and lists them, along with credit appraisal scores, in the DMI file (DUNS Marketing Information). Because suppliers and financial institutions use the DMI to establish lines of credit, establishments have an incentive to participate and the response rate is high.

Neumark et al. (2007) compare the California NETS extract to other employment data sources, such as the Quarterly Census of Employment and Wages (QCEW). They find a high correspondence of NETS and QCEW employment levels, by county and industry, as well as employment changes over several years, also by county and industry. The main difference in employment totals between the sources is that NETS includes the self-employed in establishment employment totals, and thus also includes a large number of small businesses with no employees other than the owner or owners.

The present analysis takes advantage of several features of NETS data. First, because

DUNS numbers track establishments from year to year, a new establishment's birth year may be identified.[17] NETS compiles a longitudinal database from the Dun & Bradstreet records and also estimates missing values and cleans up irregularities in the time series data. Second, because establishments provide addresses, their location may be geocoded with precision to latitude and longitude coordinates. Geocoded establishments are then matched to year 2000 census tracts, which are typically smaller than ZIP codes (California has more than 7,000 census tracts and fewer than 2,000 ZIP codes).[18] This precision allows for a detailed level of spatial analysis and allows for direct matching with demographic data. Although any base year could be used, 2000 is selected in order to match with Decennial Census worker data.

Total new establishment employment, $NewEmp$, is also from NETS and covers four years subsequent to the base year. The measure includes establishments where NETS first reports employment from 2002 to 2005. Employment totals are as of 2005. The four year period for new establishments helps to ensure that a representative sample of births is examined, as an individual year might be idiosyncratic. Furthermore, because the likelihood of Dun & Bradstreet identifying an establishment increases with the number of years that establishment has existed, a cohort of establishments with an earlier birth year will be more comprehensive of the actual cohort founded in that year, than a cohort with an later birth year would be. Thus, using a range of birth years that includes earlier years makes the sample more comprehensive. The year 2001 is not included based on the assumption that new establishments' location choices lag their observation of the spatial distribution of employers and workers by a short period. Lastly, as discussed

[17]The database also tracks establishment relocations, which may also respond to productivity advantages but are not studied here.

[18]Geocode data and census tract matches are courtesy of the Public Policy Institute of California. Neumark and Kolko (2008) describe the geocoding process, which used ArcGIS mapping software to match NETS addresses by street name, number, and ZIP code to a street map database. By using two rounds of geocoding and varying the level of sensitivity to establish matches, as well as other methods, they were able to successfully geocode over 95 percent of establishment-year observations.

in the methodology section, $NewEmp$ only includes new establishments with with > 1 employee and ≤ 500 employees.

While the employment variables aggregate establishment level data to the census tract level, worker data is only available at the census tract level. The year 2000 Decennial Census long form tabulations, derived from responses of a 1 in 6 sample of households, give estimates of the totals that would have been obtained from a complete count. It is crucial to the present analysis that these totals are by place-of-residence census tract. Among the economic variables included is the primary occupation of those 16 or older that are employed, or unemployed for less than 5 years. From this variable, the present analysis defines for each industry: total workers specialized in an occupation crucial to that industry, $WrkSpec$, and the total of all other workers, $WrkOther$.

Table 2 gives statewide totals for the census tract level variables. Total new establishment employment represents from 8% to 22% of each industry's base year employment ($EmpSame$). Comparing employment and worker totals, there are about 1.5 million more employees, counted by place-of-work, than workers, counted by place-of-residence. One explanation of this discrepancy is that some workers hold multiple positions, and could be recorded as employees of, or self-employed in, multiple businesses in the NETS data (which includes part-time employees). Thus, the number of jobs could be greater than the total number of workers as measured by the Decennial Census, which for this variable counts workers only by their primary occupation. Because the combined industries account for only 3.1% of California employment and the combined specialized occupations, account for only 9.2% of California workers, the remaining totals, $EmpOther$ and $WrkOther$, are similar across industries. Lastly, note that there are fewer employees in each industry than there are workers with specialized occupations crucial to those industries. This difference suggests that the occupation definitions are not overly specific to each industry, and that a substantial portion of the specialized workers each industry is

attracted to may be employed in other industries in the base year.

4.3 Access variables

Because the term for locational characteristics, y_j in Equation (1), may reflect distinct benefits originating from employment within the same census tract as a new establishment and employment in surrounding census tracts, two sets of access variables are defined.

The first set of four variables measure in-tract access to employers and workers. Each variable gives the the level of employment located in, or workers residing in, tract j as a share of all employment or workers in California. The in-tract variables are written as

$$EmpSameIn_j = \frac{EmpSame_j}{\sum_{i=1}^{J} EmpSame_i}, \quad EmpOtherIn_j = \frac{EmpOther_j}{\sum_{i=1}^{J} EmpOther_i}, \tag{2}$$

$$WrkSpecIn_j = \frac{WrkSpec_j}{\sum_{i=1}^{J} WrkSpec_i}, \quad WrkOtherIn_j = \frac{WrkOther_j}{\sum_{i=1}^{J} WrkOther_i}. \tag{3}$$

Constructing the variables as shares makes the distribution of access to different sources directly comparable across industries and sources. For example, sources with greater standard deviations in in-tract access are more spatially concentrated statewide. The access measures are not normalized for census tract size.[19]

A second set of four variables measure access to employers or workers outside of census tract j, as a distance weighted sum of the statewide shares in all surrounding census tracts, within 20 miles. For the weighting, d_{ij} is the distance between the centroids (the spatial centers) of census tracts i and j (calculated in ArcGIS). The outside of tract

[19] An alternate formulation of in-tract access would divide in-tract totals by the tract's radius to measure that less compact tracts may not have as good of access to all employment and workers located in the tract (Graham, 2007). One drawback to the radius method is that in many tracts, employment may actually be concentrated in a specific area, so productivity spillovers are not diminished.

variables are written as

$$EmpSameOut_j = \sum_{i=1}^{J} \frac{EmpSameIn_i}{d_{ij}}, \quad EmpOtherOut_j = \sum_{i=1}^{J} \frac{EmpOtherIn_i}{d_{ij}}, \quad (4)$$

$$WrkSpecOut_j = \sum_{i=1}^{J} \frac{WrkSpecIn_i}{d_{ij}}, \quad WrkOtherOut_j = \sum_{i=1}^{J} \frac{WrkOtherIn_i}{d_{ij}}, \quad (5)$$

where $i \neq j$ and $d_{ij} < 20$. This weighting scheme assumes that the density of nearby census tracts is more important for agglomeration than that of distant census tracts.[20] Distances are cut off after 20 miles because MSA fixed effects are meant to account for regional differences in productivity that vary only on a larger spatial scale.

The presence of in-tract employment, as opposed to out-of-tract employment, is especially relevant for new establishment location choice. Equations (3) assume that establishments located in a census tract have equal access to all other establishments in that tract. This assumption is based on the design rule that all addresses in a census tract must be accessible from within the tract (implying convenient access and interaction), and that even if a census tract is spatially large, workplaces will tend to be more spatially concentrated. Many urban census tracts are less than a mile across. Because of this immediate access, actually locating in a high employment census tract may facilitate especially high productivity-enhancing interaction, such as better communication and knowledge spillovers as well as the presence of complementary businesses in other industries. Rosenthal and Strange (2003) find that existing employment within one mile of a ZIP code centroid has a much greater effect on new establishment employment than

[20] These variables definitions offer a middle road of spatial detail compared to previous intra-urban analyses. The present analysis takes the attenuation of economies of agglomeration for granted, and so does not measure how benefits decline with distance as in Rosenthal and Strange (2003), where access is measured in rings of 1 mile, 1 to 5 miles, 5 to 10 miles, and 10 to 15 miles from each ZIP code centroid. However, the present analysis does not simplify spatial data as much as Graham (2007), where access is measured as a single, distance weighted sum of employment, with no distinction of employment in a jurisdiction from employment surrounding a jurisdiction.

existing employment at even moderately further distances. Aharonson et al. (2007) find an especially large spillover between biotechnology businesses in the range of only 500 meters.

In addition to measuring local productivity spillovers, the extent of in-tract employment also indicates whether there is commercial or industrial real estate available for new establishments to locate in. To exclude some of the lowest employment census tracts where there are likely to be few location options, the sample only includes tracts with more than a 0.00001 share of total statewide employment.[21] This threshold reduces the sample from 7046 to $J = 6668$ census tracts. Note that even though some census tracts are dropped as observations, the access to out-of-tract employment and workers independent variables still include them in distance weighted totals.

Table 3, with summary statistics for new establishment employment and all the access variables, shows the higher variance of census tract employment data, compared to worker data. The main reason for this difference is that census tracts are designed based on residential population, not place-of-work employment data. Therefore, while the number of residents is fairly constant across census tracts (approximately 4000 residents), total employment has high variance. For each industry, the highest in-tract variance is for same-industry employment, followed by other-industry employment, specialized workers, and other workers. Variance of out-of-tract employment access is also greater than out-of-tract worker access.

Other statistics in Table 3 also characterize the spatial concentration of new establishment employment and the eight access variables. The maximum new establishment employment in any tract is 2.1% to 6.7% of total new establishment employment, depending on the industry. Three industries have positive new establishment employment in over 1000 census tracts. The Instrument industry, which is more concentrated in

[21] Section 5.4 presents results for even more restrictive samples using higher employment thresholds.

manufacturing areas, has new establishments in only 306 census tracts.

Regarding the access variables, some census tracts do account for a major share of certain industry's statewide employment. The Movie and R & D industries have the greatest variation in share of employment, indicating that they are more concentrated. 15% of Movie industry employment is in a single census tract and 13% of R & D employment is in a single census tract. In contrast, the maximum census tract for other industry employment has less than 1% of statewide other industry employment. Worker shares are even more diffuse.

In addition to census tract characteristics, Equation (1) also contains y_m, giving MSA level fixed effects (discussed in Section 3). The present analysis for California includes MSA fixed effects for the 10 largest MSAs (or groups of MSAs), with over 100 census tracts each.[22]

4.4 Correlations across access variables

Because variation in the distribution of employers and workers provides identification in the regression model, it is important to examine correlations between the independent variables. With high correlations, as is common with spatial data, multicollinearity would make it difficult to infer the effect of a change in one variable, conditional on the other variables. Table 4 presents partial correlations of in-tract access and out-of-tract access (referred to as correlations hereafter), with MSA effects removed as in the regressions. For each industry, the correlations in the upper left quadrant of the table are between in-tract variables, those in the lower left are between in-tract and out-of-tract variables,

[22]MSAs are defined by counties and the fixed effects include: Los Angeles/Orange counties (2629 tracts), Alameda/Contra Costa/Marin/San Francisco/San Mateo counties (871 tracts), San Diego county (604 tracts), Riverside/San Bernardino counties (587 tracts), San Benito/Santa Clara counties (349 tracts), Fresno county (158 tracts), Ventura county (155 tracts), Kern county (140 tracts), El Dorado/Placer/Sacramento/Yolo counties (124 tracts), San Joaquin county (121 tracts). The results change little if Los Angeles/Orange counties are separated.

and those in the lower right are between out-of-tract variables.

Almost all correlations between in-tract access variables are low, except for two above 0.50. The highest correlations are for $EmpSameIn$ with $EmpOtherIn$ and for $WrkSpecIn$ with $WrkOtherIn$, simply indicating that where there is a large share of total employment or population, there tends to also be a larger share of industry or worker groups.

Likewise, almost all correlations of in-tract access with out-of-tract access are low; the highest being $WrkSpecIn$ with either $EmpSameOut$ or $WrkSpecOut$. Tracts near high employment for an industry or near high residency of a specialized worker occupation tend to have a higher share of those workers. The correlations are much lower for within tract employment ($EmpSameIn$) because some tracts have little or no employment, even in dense areas.

In contrast, all correlations between out-of-tract access variables are high. The highest correlation for each industry (0.93) is for $EmpOtherOut$ with $WrkOtherOut$, reflecting an overall alignment of employment and residential density gradients in urban areas. The second highest correlation (ranging from 0.63 to 0.75) is for $EmpSameOut$ with $WrkSpecOut$, reflecting co-agglomeration of industries with specialized workers.

These high correlations could complicate attributing productivity advantages to benefits from access to employment or access to workers. However, there is a theoretical basis for excluding some of the variables from regressions where others are present. First, while other-industry employment may have higher variance and decline from a city-center with a steeper gradient than other-occupation workers, both sets of "other" access variables proxy for the overall benefits and costs associated with the local density of an urban area. Therefore, $WrkOtherIn$ and $WrkOtherOut$ may be excluded without a major loss of information on productivity benefits and costs. Second, while access to in-tract same-industry employers offers distinct benefits compared to out-of-tract employers (as

discussed above), there is less reason to distinguish between in-tract and out-of-tract specialized workers. The benefit of proximity to workers depends primarily on commute cost, determined by distance or travel time, so there is unlikely to be a discontinuous benefit from locating in the same tract as workers compared to locating nearby them. The in-tract and out-of-tract specialized worker variables, are combined as

$$WrkSpecAll_j \;=\; WrkSpecIn_j + WrkSpecOut_j. \tag{6}$$

Table 5 presents summary statistics for $WrkSpecAll$ as well as partial correlations with the access to employers variables. The combined variable's mean, variance, and correlations are similar to those of $WrkSpecOut$ (including in-tract workers with a distance weighting of one makes little difference). Regressions including access to out-of-tract employment and workers will use this reduced variable set.

Figures 1 to 4 map the access variables by census tract, with the darker shades indicating greater in or out-of-tract access. All the figures are for Software industry variables and depict the San Francisco Bay area. Figure 1, for $EmpOtherOut$, shows how access to other industry employment is distributed, and indicates that census tracts in the financial district of San Francisco have access to the greater share of statewide employment than anywhere else in the area. Figure 2, for $EmpSameIn$, shows in-tract Software industry employment. Even though some of the high employment tracts are very large, most of the land area in these tracts is uninhabited hills or wetlands and the employment centers within the tracts are actually fairly concentrated. Figure 3, for $EmpSameOut$, shows access to out-of-tract software industry employment, with the greatest densities in downtown San Francisco as well as Silicon Valley. Note how the out of out-of-tract measure smooths access compared to Figure 2. Lastly, Figure 4, for $WrkSpecAll$, shows access to workers who are specialized in computer and mathematical

occupations. Although these workers are also heavily concentrated in the San Francisco and Silicon Valley areas, the best access to them is somewhat displaced from the core of the employment centers.

5 Results

This section presents regression results for four specifications and examines a few alternatives including: a travel time based measure of access, spillovers from other industries that employ the same occupation of workers, a sample limited to higher employment census tracts, and an analysis of the importance of spatial autocorrelation.

5.1 Results with access measured by distance

For each industry, Table 6 reports coefficients and significance levels for four regression specifications, with the main result in specification 4. Although many of the coefficients are very large in magnitude, they correspond with the very small share variables. To make the magnitudes more intuitive, and also to give a sense of how productivity benefits may vary with a tract's access to employment and workers, Table 6 reports the both the marginal effects on the number of new establishment employees and the effect of a one standard deviation increase in the access variables from their mean values.

The first specification, labeled 1, only reports the effects of in-tract shares of employment or workers on new establishment employment. Even though it is expected that out-of-tract access will also provide agglomeration benefits, in-tract access alone has a positive and significant effect for the first three variables, $EmpSameIn$, $EmpOtherIn$, and $WrkSpecIn$, for each industry. In this and other specifications, most of the coefficients are significant at the 1% or 5% levels.

The second and third specifications report benefits of agglomeration based on access

to either workers or employment, in and out-of-tract. The pattern of results is similar across both specifications (substituting access to specialized workers for access to same-industry employment and access to other-industry workers for access to other industry employment). All the estimates are significant with the exception of the Instrument industry, where in-tract other-occupation workers, $WrkOtherIn$, is not significant. One reason for the insignificance may be that Instrument industry establishments need to locate in manufacturing areas, so a large residential population could limit available real estate. Nevertheless, because most of the employment and worker access measures are correlated, these separate regressions cannot establish the distinct contributions of each source to economies of agglomeration, as is done in specification 4.

Note that the employment results in specification 3 follow the findings of previous spatial analyses of agglomeration effects (Rosenthal and Strange, 2003).[23] First, greater in-tract or out-of-tract access to same-industry employers, also known as "economies of localization," induces more new establishment employment to locate there. Second, census tracts with more in-tract other-industry employment have greater economies of agglomeration. Third, the effect of access to out-of-tract other-industry employers is negative. Theoretically, the access to other-industry employers effect, also known as "economies of urbanization," may be positive or negative depending on whether an industry is better or worse off being in an especially dense part of an urban area. The positive effect of in-tract other-industry employment may reflect the abundance of real estate options and convenient access to business services associated with high density. While access to out-of-tract other-industry employment could also provide benefits, the

[23]To understand the magnitude of the coefficients, consider the effect of an increase in access to employment of one standard deviation above the mean, as presented in the right four columns of Table 6. For specification 1, increasing any of the first three variables by one standard deviation increases new establishment employment by 5 to 17 employees. However, there is no clear pattern across the industries of which factor is most important. In specifications 2 and 3, a standard deviation increase in the first three variables of each specification also attracts a similar number of additional jobs.

negative effects of density, including high rental rates for commercial and industrial real estate and other congestion costs, appear to dominate.

Specification 4 resolves the inconclusiveness of regressions 2 and 3 by including variables for both access to employment and access to workers, giving several results. First, while the coefficients for access to employment retain their signs and significance, $WrkSpecAll$, also has a positive and significant effect on new establishment employment. Second, for each industry, the effect of access to out-of-tract same-industry employers, $EmpSameOut$, falls in magnitude. Third, for each industry, the in-tract effects of same-industry and other-industry employment are stable in magnitude. Fourth, the effect of $EmpOtherOut$ becomes more negative.

These results suggest that access to specialized workers, by place of residence, accounts for a portion of the productivity effect attributed to access to out-of-tract same-industry employers in its absence. Specification 3, along with other spatial analyses of productivity effects such as Rosenthal and Strange (2003), implicitly assume that all local benefits of agglomeration derive from proximity to other employers. Rosenthal and Strange (2003) suggest that a portion of the estimated benefit could be from labor market pooling, because access to employers would also indicate access to a thick labor market. Table 4 shows that indeed, access to same industry employers and specialized employees is positively correlated. Specification 4 adds the finding that access to specialized workers also provides a distinct benefit to new establishments.[24]

Analyzing the magnitudes of the coefficients in specification 4 suggests that the effect of labor market pooling may be large compared to employer productivity spillovers. For most industries, the coefficient of access to specialized workers is substantially larger

[24]This result is consistent with the finding from several inter-industry analyses that worker specialization is a major driver of an industry's spatial concentration, relative to other industries (see Section 2).

than the coefficient of access to out-of-tract same-industry employment.[25] Furthermore, in Table 6, the effect of a standard deviation increase in access to workers is greater than that for access to same-industry employers in the Software, Movie, and R & D industries. One exception is the Instrument industry, where the magnitude of the worker effect is more similar to that of *EmpSameOut*, and the coefficient is insignificant. This smaller effect and insignificance may be due to engineering occupations being a poor measure of specialized workers for the Instrument industry.

5.2 Results with access measured by travel time

The analysis above uses distance between census tract centroids, d_{ij}, to weight access to employers and workers. However, because some productivity benefits require interaction, such as face to face meetings and service or product deliveries, and because most workers must commute to their employers, travel time may be a more accurate measure of access. As evidence of this point, Graham (2007) finds that travel time weighted access to employers is a better measure of productivity spillovers, between wards in the UK, than distance weighted access. To the extent that the indirect layout of road networks and urban traffic congestion increase travel time relative to distance, distance weighted access will be a poor measure of agglomeration potential.

This subsection repeats the last specification for a sample of the data where travel time between census tracts is available. While NETS and Census data are available for all of California, adequate travel time data were only obtained from selected Metropolitan Planning Organizations (MPOs) including: The Southern California Association of Governments, or SCAG (comprising Los Angeles, Orange, Ventura, San Bernardino,

[25]Note again that the descriptive statistics of the variable *WrkSpecAll* are similar to *WrkSpecOut*, changing little with the addition of in-tract workers. Thus, the scales of *WrkSpecAll* and *EmpSameOut* are directly comparable (with similar means), both based primarily on distance weighted access to statewide shares of each variable.

Riverside, and Imperial Counties) and the San Diego Association of Governments, or SANDAG (San Diego County). The data for both MPOs gives a complete matrix of simulated, drive alone, travel times in the morning peak period between all Transportation Analysis Zones (TAZs), as both origins and destinations, for each jurisdiction. The simulations assume trip generation and infrastructure capacity for the base year of 2003. The time weighted access measures, using t_{ij} for travel time instead of d_{ij} for distance, are otherwise identical to equation (5), except that shares are calculated only for southern California.[26]

Table 7 presents regression results based on travel time for each industry. Because the sample includes only southern California census tracts, the coefficients are not directly comparable to those in Table 6. Instead, Table 7 presents regressions for both distance and travel time side by side, using only the southern California sample. For brevity, Table 7 only presents specification 4 results, using the four industry variables along with specialized workers.

For each industry and variable, the estimates have the same signs as in the distance weighted regressions, and most are also significant. Note first that the coefficients for the in-tract variables, which do not depend on travel time, are very similar between the distance and travel time regressions. However, because the distance and travel time variables are measured on different scales, it is impossible to compare the coefficients directly.

Instead, consider the relative changes in the magnitudes of effects of access to employment and access to workers variables. The relative shifts in the magnitudes of effects between the distance and travel time regressions suggest that benefits from access to

[26]Both MPOs provide matches of TAZs to Census tracts. Because multiple TAZs may correspond to a single Census tract, t_{ij} averages travel times between all TAZs corresponding to each pair of Census tracts, for each flow direction. The 20 mile limit still applies and access measures assume no access between SCAG and SANDAG census tracts, which are in fact separated by sparsely or unpopulated areas along most of the border (as of the early 2000s).

specialized workers are especially sensitive to variation in travel time. For each of the first three industries' travel time specifications, the magnitude of a standard deviation increase in $WrkSpecAll$ grows relative to the effect of $EmpSameOut$, compared to the distance specification. Once again, the access to workers variable for the Instrument industry responds little, relative to access to same-industry employment.

One interpretation for the larger effect of access to workers in the travel time specifications is that workers, most of whom commute from home to an establishment on a daily basis, are especially averse to traffic congestion. Because the measure of travel times is directional, it reflects the difficulty of assembling workers at a census tract during the morning peak, and the difficulty of dispersing those workers back to their homes during the afternoon peak. In contrast, businesses may be able to schedule interactions at times with less congestion, having meetings in the middle of the day and receiving or making deliveries at night. Thus, because workers are less flexible in terms of commuting access, new establishments may find it especially important to locate in locations accessible to workers in order to benefit from labor market pooling.

5.3 Economies of agglomeration across industries

The previous results sections attributed the benefit of access to workers, from the variable $WrkSpecAll$, to labor market pooling. An alternate explanation for the positive effect of $WrkSpecAll$ on new establishment employment is that other nearby industries, which also employ those workers, may themselves be responsible for economies of agglomeration. These effects could include productivity spillovers across industries or benefits related to economies of scale in input production. For example, software industry establishments may benefit from proximity to computer manufacturers, which also employ workers with mathematical and computer related occupations. Even though the access to workers

variable is based on residential location, whereas such spillovers would presumably occur between workplaces, it is possible that such relationships bias the estimate of the effect of $WrkSpecAll$ upward.

One means of examining the magnitude of this bias is to include a variable for access to the places-of-work of the specialized workers, as opposed to their places of residence. The workplace density of these workers should control for any productivity spillovers originating from the industries they are employed in that might otherwise be interpreted as benefits of labor market pooling. Census tract level employment counts by place-of-work and occupation are available from the Census Transportation Planning Package (CTPP), a special tabulation from the Decennial Census long form.[27] Access to these workers is calculated in the same manner as access to same-industry employment, $EmpSameOut$, and is denoted $WrkPlaceSpecOut$.[28]

Table 8 presents specifications including workers by place-of-work, $WrkPlaceSpecOut$, as well as the employment and place-of-residence variables. Specification 1 omits access to workers by place-of-residence, and finds that access to workers by place-of-work does have a significant and positive effect on new establishment employment for three industries, and is also positive for the Software industry. In specification 2, $WrkPlaceSpecOut$ continues to have the expected sign of a positive effect on new establishment employment, but the effect is only significant for the R & D industry (possibly indicating the importance of universities). Compared with specification 4 of Table 6, the magnitude of the $WrkSpecOut$ effect is lower for each industry, indicating that there may have been

[27]The CTPP consists of three parts, tabulating responses to both transportation related questions as well as other Decennial Census topics by respondent's place-of-residence, place-of-work, and flows between places of residence and work places. Tabulations are aggregated to the census tract level. The Census Bureau determines places of work based on respondent provided addresses, followed by various imputation procedures in cases where the address is missing or faulty.

[28]This measure does potentially duplicate the measures of same industry employment for especially concentrated industries. For example, about one quarter of workers in computer and mathematical occupations are employed in the Software industry.

a positive bias. However, the effect of access to workers by place of residence is still significant where it was before, and the magnitude is still substantial relative to the other effects. For all except the Movie industry, *EmpSameOut* also decreases in magnitude. Thus, benefits from related industries may also have biased *EmpSameOut* upward.

5.4 Minimum employment threshold

Another concern with the estimates above is that not all of the approximately 7046 census tracts in California, or even the 6888 used in the above results, may be suitable for new establishments to locate in. In particular, because census tracts are designed around place of residence population and not around places of work, some may be zoned predominantly for residential use (as is evident from the high variance of in-tract employment in the summary statistics). Census tracts with little or no in-tract employment may lack the commercial or industrial real estate necessary for new establishments to locate there, regardless of whether there is good access to out-of-tract employers.[29]

Table 9 presents regression results for a sample of census tracts that have other-industry employment greater than specified thresholds of: a 0.00005 share of statewide employment, giving $J = 4125$ census tracts, and an even higher 0.0001 share, giving just $J = 2431$ census tracts. These thresholds exclude census tracts with employment levels significantly lower than the typical residential population of a census tract. Census tracts surpassing this threshold of about 1,600 employees (for the 0.0001 share) should have sufficient business activity to offer suitable location opportunities for new establishments.[30]

Compared with the full sample results in Table 6 the effect of access to all measures

[29]Rosenthal and Strange (2003), use ZIP codes, a larger spatial unit, where there are likely to be more opportunities for new businesses to locate. They do not specifically measure access to employment within the ZIP code; rather, their smallest range is employment within 1 mile of a ZIP code centroid, which may include portions of other ZIP codes (assuming that employment is spread evenly throughout ZIP codes).

[30]Even an exclusively residential census tract could have positive employment due to the self-employed using a home address for their business.

of employment are relatively more important than access to workers, but significance levels remain unchanged despite the smaller sample size. One reason for this shift in magnitude may be that in more urban areas there is even more variance in access to employment than in access to workers, so employment has greater explanatory power. Nevertheless, access to workers, $WrkSpecOut$, still has a significant effect that is of comparable magnitude to $EmpSameOut$.

5.5 Spatial autocorrelation

The present analysis, like most of the papers cited above, has so far assumed that outcomes in neighboring census tracts are independent of one another. However, errors in nearby census tracts may be correlated due to unobserved, local factors that affect new establishment employment. If spatial autocorrelation is present, standard errors may be understated and some of the coefficients deemed significant in the above regressions may be insignificant.

While a spatial tobit, which uses a spatial weighting matrix on the error term, would control for spatial autocorrelation, such a model is complex to implement and requires assumptions on the spatial weighting.[31] As an alternative, the present analysis uses a combination of clustering standard errors and spatially isolating census tracts from one another to reduce the potential for spatial autocorrelation. First, in Table 10, specification 1 repeats the primary specification (Table 6 specification 4), but with errors clustered at the MSA level. Clustering assumes that the errors within MSAs are not independent and would be expected to increase standard errors. While the significance level does fall on some results, most of the access to employers and workers variables that were significant before remain significant.

[31] One option is to use a Bayesian spatial model. LeSage (2000) proposes a Gibbs sampler to estimate spatial, limited dependent variable models, including the tobit.

Second, in specification 2, which also clusters standard errors by MSA, the sample is reduced so that the remaining census tracts are more spatially isolated from one another. This spatial isolation is achieved by selecting at most one census tract from each Zip Code Tabulation Area (ZCTA). The Census Bureau creates ZCTAs to approximate the coverage of one or more 5 digit Zip Codes. There are 1750 ZCTAs in California compared with 7046 census tracts. With only 6668 census tracts having at least a 0.00001 share of statewide employment and being in use, there are only 1255 ZCTAs with one or more census tracts available.[32] This method contrasts from that of Section 5.4, where the sample was also reduced, in that selection by ZCTAs reduces the likelihood that included census tracts will be close to one another.

Specification 2 chooses the census tract with the median employment level of all tracts in the ZCTA. The robustness of the results to spatially isolating census tracts varies somewhat by industry. The effect of out-of-tract access to same-industry employers becomes substantially less than that of access to workers. In both the Software and Movie industries, the same-industry employment effect becomes insignificant while the worker effect is still significant at the 5 percent level. In both the R & D industry and Instrument industry the out-of-tract effects are insignificant. From these results, it seems that in the industries where the agglomeration effects have consistently been the most strong and significant, there is still an effect of access to workers on new establishment employment.

6 Conclusion

In conclusion, this analysis estimates the effect of access to employers and access to workers on new establishment employment and finds that access to workers and access to

[32]Because Zip Codes and census tracts are not designed to share boundaries, there is no standard mapping of census tracts to Zip Codes. The Census Bureau does release ZCTA shape files, so census tract centroids were matched to ZCTAs using ArcGIS software.

employers are correlated, but that both sources contribute to economies of agglomeration for most of the industries examined. While previous analyses measured agglomeration effects originating from proximity to employers, the present analysis shows that for some industries, access to specialized workers may have a positive effect as well. Furthermore, it shows that access to workers is especially sensitive to travel time access, suggesting that traffic congestion may limit the extent of labor market pooling within urban areas.

Future work on this topic could make greater use of the year to year tracking of establishments in the data set. A longitudinal analysis could examine how establishment deaths contribute to establishment births within urban areas. Furthermore, it may be important to measure the location choices of workers along with employers, as both groups have an incentive to move towards one another.

7 References

Anas, A., Arnott, R., Small, K., 1998. Urban Spatial Structure. Journal of Economic Literature 36, 1426-1464.

Aharonson, B., Baum, J., Feldman, M., 2007. Desperately seeking spillovers? Increasing returns, industrial organization and the location of new entrants in geographic and technological space. Industrial and Corporate Change 16, 89-130.

Andersson, F., Burgess, S., Lane, J., 2007. Cities, matching and the productivity gains of agglomeration. Journal of Urban Economics 61, 112-128.

Audretsch, D., Feldman, M., 1996. R & D Spillovers and the Geography of Innovation and Production. The American Economic Review 86, 630-640.

Bacolod, M., Blum, B., Strange, W., 2008. Skills in the City. Journal of Urban Economics (forthcoming).

Bleakley, H., Lin, J., 2007. Thick-market effects and churning in the labor market: Evidence from U.S. cities. Federal Reserve Bank of Philadelphia, Working Paper, No. 07-23.

Chinitz, B., 1961. Contrasts in agglomeration: New York and Pittsburgh. American Economic Review 51, 279-289.

Crane, R., 1996. The Influence of Uncertain Job Location on Urban Form and the Journey to Work. Journal of Urban Economics 39, 342-356.

Davis, S., Haltiwanger, J., Jarmin, R., Krizan, C.J., Miranda, J., Nucci, A., Sandusky, K., 2006. Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes. NBER Working Paper 13226.

Dumais, G., Ellison, G., Glaeser, E., 1997. Geographic concentration as a dynamic process, NBER Working Paper 6270.

Duranton, G., Puga, D., 2004. "Micro-foundations of urban agglomeration economies," in J.V. Henderson and J.-F. Thisse, eds., Handbook of Urban and Regional Economics, Volume 4 (New York: North Holland).

Ellison, G., Glaeser, E., 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. Journal of Political Economy 105, 889-927.

Ellison, G., Glaeser, E., Kerr, W., 2007. What causes industry agglomeration? Evidence from coagglomeration patterns. NBER Working Paper 13068.

Fallick, B., Fleischman, C., Rebitzer, J., 2006. Job-hopping in Silicon Valley: Some evidence concerning the microfoundations of a high-technology cluster. The Review of Economics and Statistics 88, 472-481.

Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? American Economic Review 98, 394-425.

Freedman, M., 2008. Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. Journal of Urban Economics 64, 590-600.

Freedman, O., Kern, C., 1997. A model of workplace and residence choice in two-worker households. Regional Science and Urban Economics 27, 241-260.

Giuliano, G., Small, K., 1993. Is the journey to Work Explained by Urban Structure. Urban Studies 20, 1485-500.

Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A., 1992. Growth in Cities. Journal of Political Economy 100, 1126-1152.

Glaeser, E., Kolko, J., Albert, S, 2001. Consumer City. Journal of Economic Geography 1, 27-50.

Graham, D., 2007. Variable returns to agglomeration and the effect of road traffic congestion. Journal of Urban Economics 62, 103-120.

Haltiwanger, J., Lane, J., Spletzer, J., 2007. Wages, productivity, and the dynamic interaction of businesses and workers. Labour Economics.

Helsley, R., 1990. Knowledge production in the CBD. Journal of Urban Economics 28, 391-403.

Helsley, R., Strange, W., 1990. Matching and agglomeration economies in a system of cities. Regional Science and Urban Economics 20, 189-212.

Henderson, J., Kuncoro, A., Turner, M., 1995. Industrial Development in Cities. Journal of Political Economy 103, 1067-1085.

Holmes, T. J. 1999. Localization of Industry and Vertical Disintegration. Review of Economics and Statistics 81, 314-25.

Jaffe, A., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. Quarterly Journal of Economics 108: 577-598.

Kim, S., 1995. Excess Commuting for Two-Worker Households in the Los Angeles Metropolitan Area. Journal of Urban Economics 38, 166-182.

Kolko, J., 2008. Urbanization, agglomeration, and co-agglomeration of service industries. NBER Conference Volume on Agglomeration (forthcoming).

LeSage, J.P., 2000. Bayesian estimation of Limited Dependent variable Spatial Autoregressive Models. Geographic Analysis 32, p. 19-35.

Ma, K., Banister, D., 2006. Excess Commuting: A Critical Review. Transport Reviews 26, 749-767.

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

Mills, E., 1972. Studies in the Structure of the Urban Economy. Johns Hopkins Press, Baltimore.

Neumark, D., Zhang, J., Wall, B., 2007. Employment Dynamics and Business Relocation: New Evidence from the National Longitudinal Establishment Time Series. Research in Labor Economics 26, 39-83.

Neumark, D., Kolko, J., 2008. Do enterprise zones create jobs? Evidence from California's enterprise sone program. NBER Working Papers 14530, National Bureau of Economic Research, Inc.

Overman, H., Puga, D., 2009. Labour Pooling As a Source of Agglomeration: An Empirical Investigation. CEPR Discussion Paper No. DP7174.

Rauch, J., 1993. Does History Matter Only When It Matters Little? The Case of City-Industry Location. The Quarterly Journal of Economics 108, 843-867.

Rosenthal, S., Strange, W., 2001. The Determinants of Agglomeration, Journal of Urban Economics 50, 191-229.

Rosenthal, S., Strange, W., 2003. Geography, industrial organization, and agglomeration. Review of Economics and Statistics 85, 377-393.

Rosenthal, S., Strange, W., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Urban and Regional Economics, vol. 4. Elsevier, Amsterdam, pp. 2119-2172.

Rosenthal, S., Strange, W., 2008. The attenuation of human capital spillovers. Journal of Urban Economics forthcoming.

Timothy, D., Wheaton, W., 2001. Intra-Urban Wage Variation, Employment Location, and Commuting Times. Journal of Urban Economics 50, 338-366.

Van Soest, D., Gerking, S., Van Oort, F., 2006. Spatial impacts of agglomeration externalities. Journal of Regional Science 46, 881-899.

Walls & Associates, NETS: National Establishment Time-Series Database, Oakland. California, 2003.

Wallsten, S., 2001. An empirical test of geographic knowledge spillovers using geographic information systems and firm-level data. Regional Science and Urban Economics 31, 571-599.

Wheeler, C. H., 2006. Cities and the growth of wages among young workers: Evidence from the NLSY. Journal of Urban Economics 60, 162-184.

Yang, J., 2008. Policy Implications of Excess Commuting: Examining the Impacts of Changes in US Metropolitan Spatial Structure. Urban Studies 45, 391-405.

Figure 1: Census tract map of the San Francisco Bay area for out-of-tract access to other-industry employment, *EmpOtherOut*, other than the Software industry. Darker shading indicates access to a higher distance-weighted share of the statewide total.

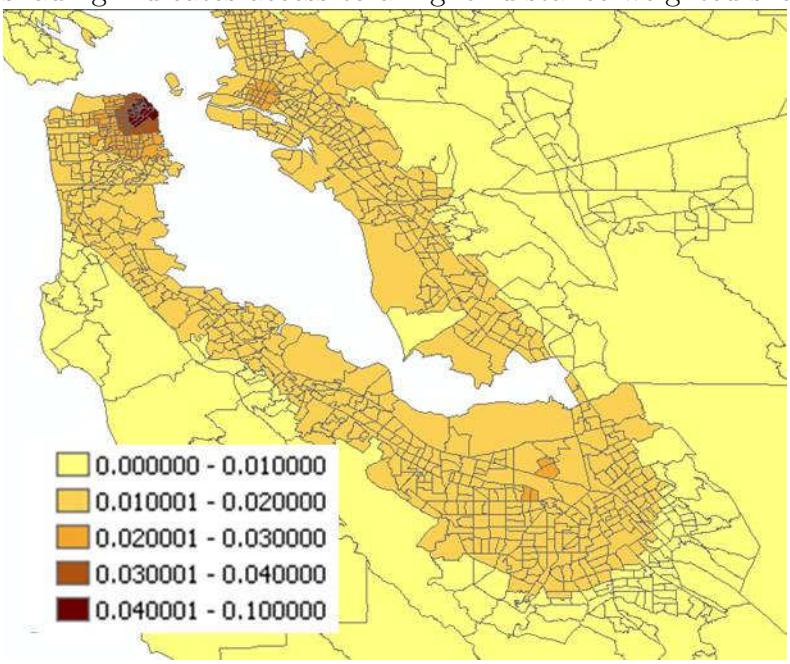

Figure 2: In-tract access to Software industry employment, *EmpSameIn*, as a share of the statewide total.

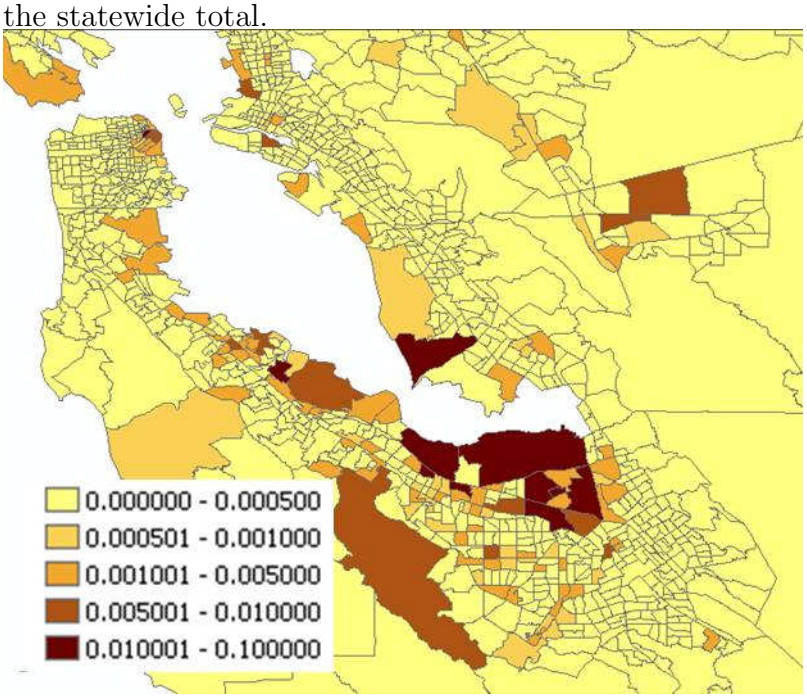

Figure 3: Out-of-tract access to Software industry employment, *EmpSameOut*, as distance-weighted shares of the statewide total.

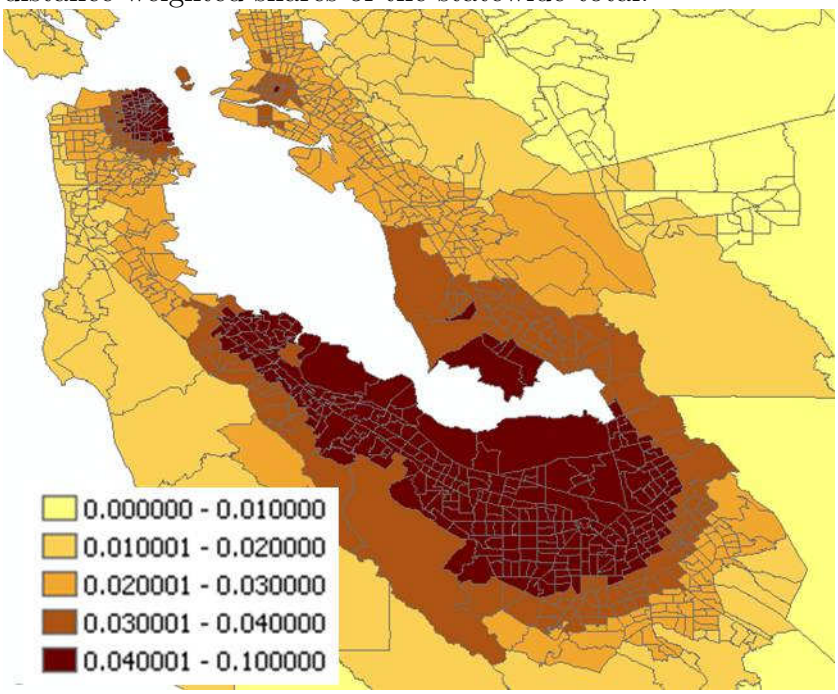

Figure 4: Access to workers in computer and mathematical occupations, $WrkSpecAll$, as distance-weighted shares of the statewide total.

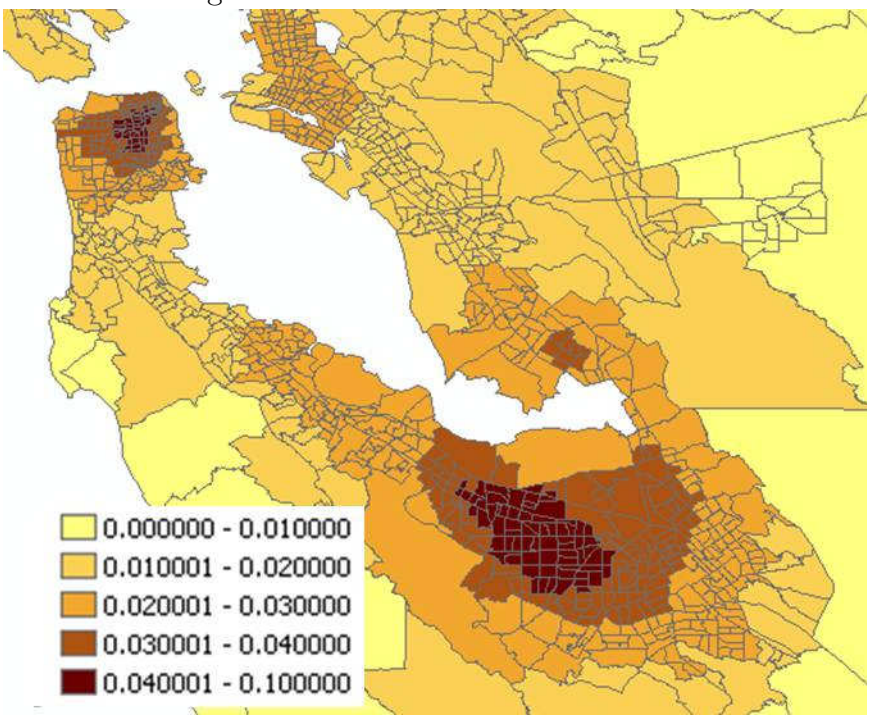

Table 1: Occupational composition of industries in California.

Occupation	Software	Movie	R & D	Instrument
Management	0.18	0.12	0.15	0.15
Business Operations	0.03	0.02	0.03	0.03
Finance	0.02	0.02	0.01	0.02
Computer and Mathematical	**0.51**	0.02	0.08	0.09
Architecture and Engineering	0.03	0.01	0.12	**0.20**
Sciences	0.01	0.00	**0.31**	0.02
Arts	0.04	**0.39**	0.02	0.02
Other Professional	0.01	0.01	0.05	0.01
Service	0.00	0.11	0.02	0.01
Sales and Office	0.13	0.20	0.15	0.18
Other (Production, Farming, Construction, ...)	0.04	0.10	0.05	0.29
Total	1.00	1.00	1.00	1.00

Employment shares are calculated from the Census 2000 Public Use Microdata Sample for California. Each of the occupations in bold is used to represent labor market thickness for the industry in which it has a large employment share.

Table 2: Totals for of census tract data for California.

Variable	Statistic	Software	Movie	R & D	Instrument
$NewEmp$	Total	28,065	25,470	11,446	8,972
$EmpSame$	Total	196,972	116,109	80,549	112,864
$EmpOther$	Total	16,038,476	16,119,339	16,154,899	16,122,584
$WrkSpec$	Total	435,939	400,742	159,546	365,006
$WrkOther$	Total	14,282,989	14,318,186	14,559,382	14,353,922

Employment data is from the National Establishment Time Series (NETS) and worker, place-of-residence data is from the Decennial Census long form. Totals include all 7046 year 2000 census tracts in California. See the variable definitions in Section 4.2.

Table 3: Summary statistics for census tract level variables.

Variable	Statistic	Software	Movie	R & D	Instrument
$NewEmp$	Mean	4.2	3.8	1.7	1.3
	Std. Dev.	22.3	24.5	14.7	17.5
	Max.	577	1174	772	524
	No. > 0	2,054	2,867	1,161	306
$EmpSameIn$	Mean	0.00015	0.00015	0.00015	0.00015
	Std. Dev.	0.00092	0.00296	0.00230	0.00153
	Max.	0.02595	0.15205	0.13086	0.05027
	No. > 0	3,790	2,433	1,631	1,191
$EmpOtherIn$	Mean	0.00015	0.00015	0.00015	0.00015
	Std. Dev.	0.00032	0.00032	0.00032	0.00032
	Max.	0.00852	0.00861	0.00858	0.00861
	No. > 0	6,668	6,668	6,668	6,668
$WrkSpecIn$	Mean	0.00014	0.00015	0.00014	0.00015
	Std. Dev.	0.00019	0.00019	0.00023	0.00017
	Max.	0.00292	0.00240	0.00574	0.00230
	No. > 0	5,848	6,183	4,893	5,949
$WrkOtherIn$	Mean	0.00014	0.00014	0.00014	0.00014
	Std. Dev.	0.00007	0.00007	0.00007	0.00007
	Max.	0.00078	0.00078	0.00078	0.00077
	No. > 0	6,652	6,652	6,652	6,652
$EmpSameOut$	Mean	0.01029	0.02180	0.00942	0.00983
	Std. Dev.	0.01353	0.04635	0.01104	0.01208
	Max.	0.12016	0.32758	0.18244	0.15203
$EmpOtherOut$	Mean	0.01190	0.01181	0.01189	0.01190
	Std. Dev.	0.01071	0.01051	0.01068	0.01069
	Max.	0.07509	0.07605	0.07563	0.07613
$WrkSpecOut$	Mean	0.01034	0.01322	0.00979	0.00981
	Std. Dev.	0.00927	0.01520	0.00862	0.00793
	Max.	0.05417	0.07825	0.04714	0.04347
$WrkOtherOut$	Mean	0.01057	0.01049	0.01058	0.01059
	Std. Dev.	0.00853	0.00833	0.00848	0.00852
	Max.	0.03401	0.03292	0.03370	0.03399

Statistics are for a sample of $J = 6668$ census tracts with total in-tract employment above a 0.00001 share of the statewide total. Variables for access to in-tract and out-of-tract employment and workers are defined in in Section 4.3.

Table 4: Partial correlations of access variables, controlling for MSA.

Industry	Variable	ESI	EOI	WSI	WOI	ESO	EOO	WSO	WOO
Software	$EmpSameIn$	1.00							
	$EmpOtherIn$	0.55	1.00						
	$WrkSpecIn$	0.18	0.13	1.00					
	$WrkOtherIn$	0.00	0.03	0.48	1.00				
	$EmpSameOut$	0.19	0.16	0.25	-0.05	1.00			
	$EmpOtherOut$	0.04	0.11	-0.02	-0.11	0.52	1.00		
	$WrkSpecOut$	0.06	0.07	0.26	-0.02	0.76	0.73	1.00	
	$WrkOtherOut$	-0.03	0.03	-0.05	-0.09	0.35	0.93	0.71	1.00
Movie	$EmpSameIn$	1.00							
	$EmpOtherIn$	0.07	1.00						
	$WrkSpecIn$	0.18	0.05	1.00					
	$WrkOtherIn$	0.02	0.03	0.35	1.00				
	$EmpSameOut$	0.14	0.03	0.44	-0.06	1.00			
	$EmpOtherOut$	0.04	0.11	0.11	-0.12	0.51	1.00		
	$WrkSpecOut$	0.12	0.04	0.44	-0.07	0.88	0.75	1.00	
	$WrkOtherOut$	0.03	0.02	0.04	-0.09	0.45	0.92	0.69	1.00
R & D	$EmpSameIn$	1.00							
	$EmpOtherIn$	0.19	1.00						
	$WrkSpecIn$	0.07	0.06	1.00					
	$WrkOtherIn$	-0.01	0.03	0.33	1.00				
	$EmpSameOut$	-0.01	0.12	0.18	-0.07	1.00			
	$EmpOtherOut$	0.00	0.11	-0.03	-0.11	0.67	1.00		
	$WrkSpecOut$	0.00	0.06	0.25	-0.06	0.73	0.74	1.00	
	$WrkOtherOut$	-0.02	0.03	-0.06	-0.09	0.56	0.93	0.71	1.00
Instrument	$EmpSameIn$	1.00							
	$EmpOtherIn$	0.39	1.00						
	$WrkSpecIn$	0.09	0.06	1.00					
	$WrkOtherIn$	0.02	0.03	0.53	1.00				
	$EmpSameOut$	0.03	0.07	0.22	0.02	1.00			
	$EmpOtherOut$	-0.02	0.11	-0.15	-0.10	0.29	1.00		
	$WrkSpecOut$	0.01	0.05	0.16	0.01	0.65	0.63	1.00	
	$WrkOtherOut$	-0.03	0.03	-0.17	-0.08	0.28	0.93	0.68	1.00

The variables in the columns are shorthand for, and in the same order as, the row variables for each industry. Correlations are partial of MSA fixed effects (see Section 4.4).

Table 5: Summary statistics for access to specialized workers.

Variable	Statistic	Software	Movie	R & D	Instrument
$WrkSpecAll$	Mean	0.01048	0.01337	0.00993	0.00996
	Std. Dev.	0.00937	0.01529	0.00869	0.00799
	Max.	0.05522	0.07958	0.04749	0.04450
	Corr w. $EmpSameIn$	0.06	0.12	0.00	0.01
	Corr w. $EmpOtherIn$	0.07	0.04	0.07	0.05
	Corr w. $EmpSameOut$	0.76	0.88	0.73	0.65
	Corr w. $EmpOtherOut$	0.72	0.74	0.73	0.62

The variable $WrkSpecAll$ gives the distance weighted sum of access to workers specialized in an occupation that is crucial to each industry. $WrkSpecAll = WrkSpecIn + WrkSpecOut$.

Table 6: Tobit regression of new establishment employment on access variables

	Estimated coefficients				Effect of 1 s.d. increase			
Industry / Variable	1	2	3	4	1	2	3	4
Software								
EmpSameIn	12206**		11896**	12017**	11.3		11.0	11.1
EmpOtherIn	28877**		30310**	30888**	9.3		9.7	9.9
WrkSpecIn	29092**	50014**			5.6	9.7		
WrkOtherIn	49982**	44021**			3.3	2.9		
EmpSameOut			640**	404**			8.7	5.5
EmpOtherOut			-782**	-1049**			-8.4	-11.2
WrkSpecOut		1388**				12.9		
WrkOtherOut		-1829**				-15.6		
WrkSpecAll				736**				6.9
Movie								
EmpSameIn	2135**		2355**	2321**	6.3		7.0	6.9
EmpOtherIn	22107**		24314**	25111**	7.2		7.9	8.1
WrkSpecIn	65679**	48009**			12.7	9.3		
WrkOtherIn	28260**	57311**			1.9	3.9		
EmpSameOut			339**	170**			15.7	7.9
EmpOtherOut			-846**	-1448**			-8.9	-15.2
WrkSpecOut		990**				15.0		
WrkOtherOut		-1630**				-13.6		
WrkSpecAll				876**				13.4
R & D								
EmpSameIn	2979**		3108**	3074**	6.8		7.1	7.1
EmpOtherIn	27604**		27587**	28667**	8.9		8.9	9.2
WrkSpecIn	23346**	21109**			5.3	4.8		
WrkOtherIn	22948**	36213**			1.5	2.4		
EmpSameOut			480**	38			5.3	0.4
EmpOtherOut			-388**	-971**			-4.1	-10.4
WrkSpecOut		1431**				12.3		
WrkOtherOut		-1255**				-10.6		
WrkSpecAll				1496**				13.0
Instrument								
EmpSameIn	4748**		4699**	4610**	7.2		7.2	7.0
EmpOtherIn	52126**		57649**	58203**	16.8		18.5	18.7
WrkSpecIn	81217**	85379**			13.6	14.3		
WrkOtherIn	-12379	-30620			-0.8	-2.0		
EmpSameOut			1622**	1236**			19.6	14.9
EmpOtherOut			-3099**	-3634**			-33.1	-38.9
WrkSpecOut		3301**				26.2		
WrkOtherOut		-4932**				-42.0		
WrkSpecAll				1492				11.9

Includes constant term and MSA fixed effects for the 10 largest urban areas in California.
Results are discussed in Section 5.1
J = 6,668 census tracts.
* Significant at 5 % level.
** Significant at 1 % level.

Table 7: Travel time measured access variables, for southern California

	Estimated coefficients		Effect of 1 s.d. increase	
Industry / Variable	Distance	Travel time	Distance	Travel time
Software				
EmpSameIn	4790**	4904**	7.6	7.8
EmpOtherIn	13460**	13419**	7.7	7.6
EmpSameOut	493**	1201**	7.7	6.2
EmpOtherOut	-808**	-2602**	-15.5	-15.6
WrkSpecAll	572**	1912**	7.5	8.2
Movie				
EmpSameIn	2102**	2168**	8.9	9.2
EmpOtherIn	20648**	20674**	11.8	11.8
EmpSameOut	297**	368*	18.0	6.4
EmpOtherOut	-1059**	-3509**	-19.9	-20.8
WrkSpecAll	302*	2364**	8.2	19.1
R & D				
EmpSameIn	3137**	3132**	13.4	13.3
EmpOtherIn	10901**	10947**	6.2	6.2
EmpSameOut	119*	41	2.1	0.3
EmpOtherOut	-632**	-2081**	-12.1	-12.5
WrkSpecAll	732**	2735**	10.3	11.7
Instrument				
EmpSameIn	1351*	1446*	3.4	3.6
EmpOtherIn	28304**	27983**	16.1	15.9
EmpSameOut	935**	2870*	11.9	12.4
EmpOtherOut	-2359**	-7431**	-45.2	-44.5
WrkSpecAll	1449**	4322**	17.2	17.9

Includes constant term and MSA fixed effects for the 4 urban areas in Southern California.
The variables *EmpSameOut*, *EmpOtherOut*, and *WrkSpecAll* are calculated with travel time weighted access. Travel time data is from the Metropolitan Planning Organizations: SCAG and SANDAG
Discussed in Section 5.2
$J = 3738$ census tracts with employment over a 0.0002 share of Southern California employment.
* Significant at 5 % level.
** Significant at 1 % level.

Table 8: Spillovers across industries sharing specialized workers.

	Estimated coefficients		Effect of s.d. increase	
Industry / Variable	1	2	1	2
Software				
EmpSameIn	11892**	12013**	11.0	11.1
EmpOtherIn	30179**	30788**	9.7	9.9
EmpSameOut	512**	315*	6.9	4.3
EmpOtherOut	-926**	-1150**	-9.9	-12.3
WrkSpecAll		729**		6.8
WrkPoWSpecOut	226	163	2.6	1.9
Movie				
EmpSameIn	2313**	2320**	6.9	6.9
EmpOtherIn	24012**	25107**	7.8	8.1
EmpSameOut	202**	169**	9.4	7.8
EmpOtherOut	-1573**	-1450**	-16.5	-15.2
WrkSpecAll		873**		13.3
WrkPoWSpecOut	660**	4	11.8	0.1
R & D				
EmpSameIn	3036**	3053**	7.0	7.0
EmpOtherIn	27468**	28223**	8.9	9.1
EmpSameOut	-45	-45	-0.5	-0.5
EmpOtherOut	-1349**	-1199**	-14.4	-12.8
WrkSpecAll		953**		8.3
WrkPoWSpecOut	1490**	685**	15.1	6.9
Instrument				
EmpSameIn	4534**	4503**	6.9	6.9
EmpOtherIn	56080**	56625**	18.0	18.2
EmpSameOut	899*	773	10.9	9.3
EmpOtherOut	-4328**	-4472**	-46.3	-47.8
WrkSpecAll		862		6.9
WrkPoWSpecOut	1871*	1618	18.5	16.0

Includes constant term and MSA fixed effects for the 10 urban areas in California. Specification includes workers in an occupation crucial to each industry by their place of residence *WrkSpecAll* and place of work *WrkPoWSpecOut*, discussed in Section 5.3

$J = 6888$

* Significant at 5 % level.

** Significant at 1 % level.

Table 9: High employment census tract samples.

	Estimated coefficients		Effect of s.d. increase	
Industry / Variable	1	2	1	2
Software				
EmpSameIn	12224**	12287**	14.2	18.4
EmpOtherIn	28024**	26594**	10.9	12.8
EmpSameOut	481**	515**	7.0	8.0
EmpOtherOut	-1024**	-1133**	-11.2	-12.7
WrkSpecAll	680**	746*	6.6	7.3
Movie				
EmpSameIn	2211**	2065**	8.3	10.1
EmpOtherIn	22947**	22478**	9.0	10.9
EmpSameOut	243**	360**	11.9	18.9
EmpOtherOut	-1357**	-1341**	-14.5	-14.8
WrkSpecAll	788**	706**	12.5	11.6
R & D				
EmpSameIn	3077**	3097**	9.0	11.7
EmpOtherIn	26940**	24740**	10.6	12.0
EmpSameOut	32	63	0.4	0.8
EmpOtherOut	-986**	-1057**	-10.7	-11.8
WrkSpecAll	1694**	1956**	14.9	17.5
Instrument				
EmpSameIn	3077**	3097**	6.0	7.8
EmpOtherIn	26940**	24740**	10.5	11.9
EmpSameOut	32	63	0.4	0.8
EmpOtherOut	-986**	-1057**	-10.7	-11.9
WrkSpecAll	1694**	1956**	13.8	15.9

The samples are defined by the share of statewide other-industry employment in each census tract as follows: Specification (1) includes only tracts with a share greater than 0.00005, giving $J = 4125$ tracts. Specification (2) includes only tracts with a share greater than 0.00010, giving $J = 2431$ tracts. Includes constant term and MSA fixed effects for the 10 urban areas in California.
Results discussed in Section 5.4.
* Significant at 5 % level.
** Significant at 1 % level.

Table 10: Spatial analysis specifications

	Estimated coefficients		Effect of s.d. increase	
Industry / Variable	1	2	1	2
Software				
EmpSameIn	12017**	11692**	11.1	10.5
EmpOtherIn	30888**	27877**	9.9	8.7
EmpSameOut	404*	100	5.5	1.3
EmpOtherOut	-1049**	-736**	-11.2	-7.6
WrkSpecAll	736*	931*	6.9	8.5
Movie				
EmpSameIn	2321**	1648**	6.9	3.2
EmpOtherIn	25111*	22968	8.1	7.2
EmpSameOut	170**	-68	7.9	-2.9
EmpOtherOut	-1448*	-1930	-15.2	-19.7
WrkSpecAll	876**	1907*	13.4	27.4
R & D				
EmpSameIn	3074	7346**	7.1	21.6
EmpOtherIn	28667**	4316	9.2	1.3
EmpSameOut	38	-77	0.4	-0.9
EmpOtherOut	-971**	-190	-10.4	-2.0
WrkSpecAll	1496**	512	13.0	4.4
Instrument				
EmpSameIn	4610**	12755**	7.0	9.0
EmpOtherIn	58203**	45508	18.7	14.3
EmpSameOut	1236**	293	14.9	3.4
EmpOtherOut	-3634**	-2589	-38.9	-26.8
WrkSpecAll	1492	1355	11.9	10.7

These specifications cluster standard errors by MSA. Specification (1) uses the $J = 6{,}668$ census tract sample. Specification (2) uses only one census tract from each Zip Code Tabulation Area (ZCTA), with $J = 1255$. Specification (2) uses only the median total employment census tract.
Results are discussed in Section 5.5
* Significant at 5 % level.
** Significant at 1 % level.